2. Chemical Property Estimation

The chemical property estimation programs in PCGEMS allow you to get a basic understanding of the nature of a chemical. Using the estimation programs, one can estimate the temperature at which the chemical melts, boils, or freezes; the rate of volatilization from water and the solubility in water of the chemical; and the fate of trace concentrations of the chemical in environmental situations. Understanding the nature of a chemical is invaluable in exposure studies.

Chemical release data must be combined with information about the chemical properties (such as water solubility) and the local environment (such as meteorological conditions or stream flow) in order to build meaningful input to models designed to estimate environmental concentrations. There are seven chemical estimation routines in PCGEMS:

- SMIGET — SMIGET gives you the structure of the chemical in SMILES notation if you have the Chemical Abstracts Number (CAS) number for it and the CAS number is found in the CAS Number/SMILES dataset. This is particularly useful since SMILES is the default structure entry method in PCGEMS.

- TITLEIII — This program gives you access to the TITLE III dataset that contains the measured and estimated values for 252 chemicals listed under Title III, Section 313 of the Superfund Amendments and Reauthorization Act (SARA).

- PCLOGP — PCLOGP estimates octanol/water partition coefficient

values.

- PCCHEM — PCCHEM estimates chemical-specific values for seven other chemical properties: water solubility, boiling point, melting point, vapor pressure, Henry's Law Constant, bioconcentration factor, and adsorption coefficient. PCCHEM also provides access to the GEMS CHEMEST dataset that contains measured values for 856 chemicals.

- PCFAP — PCFAP determines the rate of reaction and removal of a chemical in air.

- PCHYDRO — PCHYDRO determines the rate of hydrolysis of a chemical.

- DRAWSMI — This program will draw the structure of a chemical when provided with a CAS number (if that CAS number is in the CAS Number/SMILES data), a SMILES notation or an ELOGP file.

Table 2-1 lists properties that can be estimated and the models that use these properties as input.

Table 2-1. Properties that may be used as input for environmental models

PROPERTY	MODELS
Octanol/Water Partition Coefficient*	ENPART, EXAMS-II
Vapor Pressure	EXAMS-II, ENPART, ReachScan
Water Solubility	SESOIL, EXAMS-II, ENPART, and ReachScan
Henry's Law Constant	SESOIL, EXAMS-II, ENPART, and ReachScan
Carbon Adsorption Coefficient	SESOIL, EXAMS-II, ENPART, and ReachScan
Bioconcentration Factor	EXAMS-II

* The octanol/water partition coefficient is estimated by PCLOGP. All other properties are estimated by PCCHEM.

2.1 Selecting a Chemical Property Estimation Program

The menu in Figure 2-1 appears when the Chemical Property Estimation option is selected from the PCGEMS Main Menu.

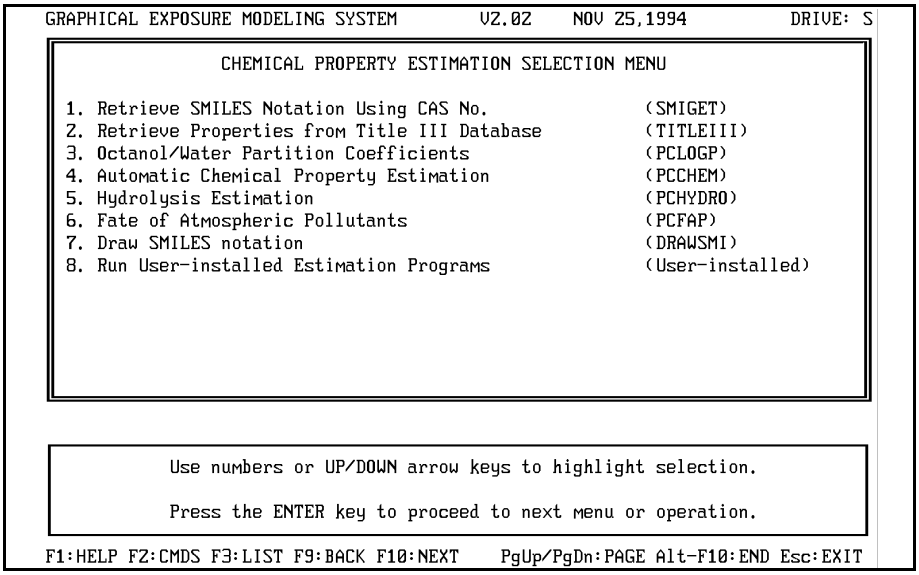

Figure 2-1. Chemical Property Estimation Selection Menu

Enter your choice for a program by positioning your cursor over the program and pressing the ENTER key. Each of the programs listed in the Chemical Property Estimation Selection Menu (Figure 2-1) is explained in detail in the following subsections.

2.2 SMILES from CAS (SMIGET)

SMILES is an acronym that stands for Simplified Molecular Interactive Line Entry System. It is a linear, chemical structure notation system that is the preferred way of describing a chemical structure for both GEMS and PCGEMS.

To make PCGEMS as "user-friendly" as possible, this program provides an automated method of obtaining SMILES notations from the CAS Number/SMILES dataset. This dataset contains the SMILES notations for over 19,000 chemicals and is helpful for those users who are not yet familiar with the SMILES nomenclature or those who wish to create an output file containing the CAS numbers, the SMILES notations, and the chemical names for a large number of chemicals.

To use this program, you must know the Chemical Abstracts Service (CAS) numbers of the chemicals. CAS numbers may be obtained from the CAS index for chemicals which is available at most libraries. In order to use this program, you must have entered the file path to the CAS Numbers/SMILES dataset in the PCGEMS configuration file.

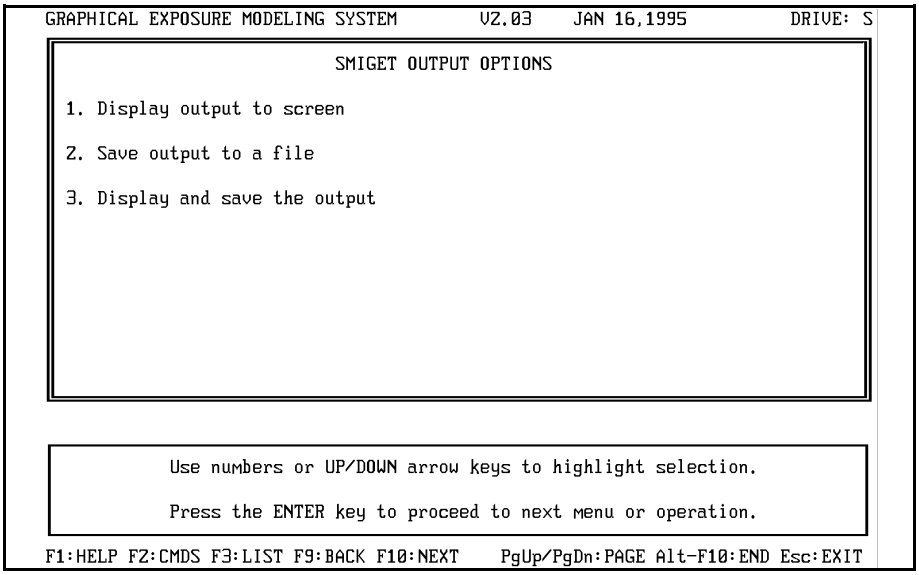

Figure 2-2. SMIGET Output Options Menu

The menu in Figure 2-2 appears when you select Option 1. SMIGET from the Chemical Property Estimation Selection Menu (Figure 2-1).

This menu offers you three options:

Option: 1. Display Output to Screen

You may choose to view the retrieved SMILES notations on your PC screen and not save them.

Option: 2. Save Output in a File

You may choose to save and not view the output on the screen. It will be written directly to a file that may be reviewed using the catalog manager.

Option: 3. Save and Display Output

You may choose to both display the model results on screen and save them.

If you choose to save the file by selecting either Option 2 or 3, the menu shown in Figure 2-3 appears, as shown below.

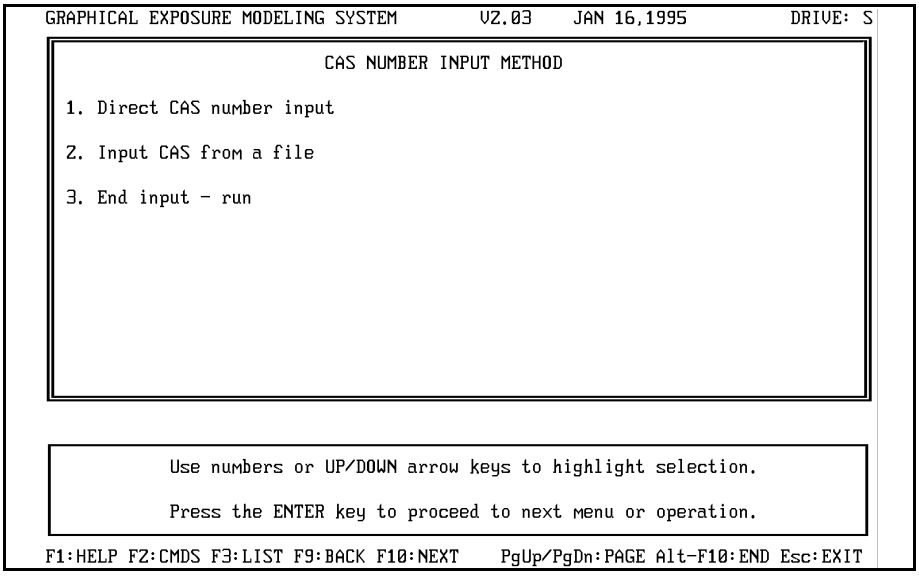

Figure 2-3. CAS Number Input Method Menu

Select an option by positioning your cursor over it and pressing the ENTER key. When you are through with the menu above, press the ENTER key. The navigational menu shown in Figure 2-4 appears. This navigational menu allows you to enter a CAS number directly or have the CAS numbers extracted from a previously created file. These entry methods are not mutually exclusive; you will be returned to this menu before the program is run. You may continue to input CAS numbers in any combination of methods. Make your selection by positioning your cursor over an option and pressing the ENTER key.

Enter the CAS number of the chemical you wish to see. You may enter it any form (with dashes, without dashes but with spaces, or without either dashes or spaces) as long as the number is valid.

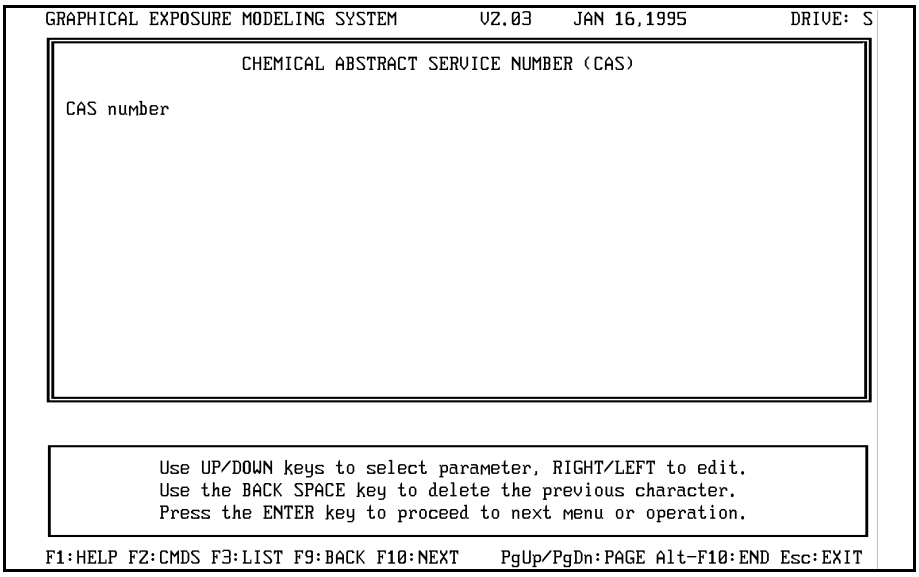

Figure 2-4. Direct Entry of CAS Number Menu

When you have entered a CAS number, press the ENTER key and PCGEMS will check to see if the CAS number you enter is valid. If the CAS number is invalid, the following message appears in the error section:

```
Invalid CAS number.
```

If the entry is correct, the menu shown in Figure 2-3 reappears, allowing you to enter more CAS numbers. You may select the direct option again, or select the file option.

If you choose the second option, "Input from a file", the menu in Figure 2-5 appears.

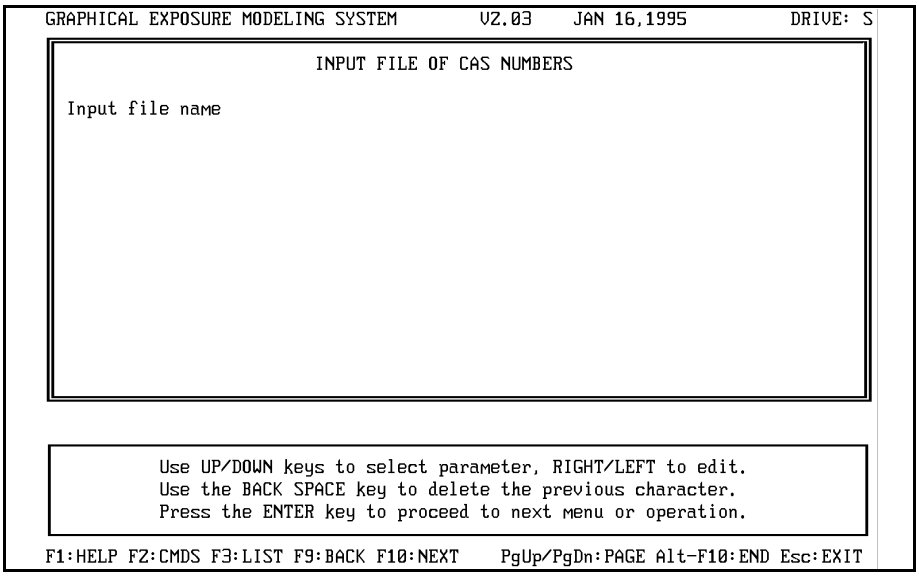

Figure 2-5. File Name Entry Menu

Enter the file where you have the CAS numbers listed. In order to use this option, the file that you enter here must follow the format provided below:

FILE FORMAT FOR CAS NUMBERS INPUT FILE:

1. The file must be in ASCII format.
2. The variable containing the CAS numbers must be the first variable in the file (columns 1-13).
3. The file extension must be "DAT".

You do not have to enter the file extension since it will be automatically added to the file name by the system. Press ENTER when you are done.

If the file is located and the format is correct, the menu shown below in Figure 2-6 appears. This menu will read your input file and show the number of records in the file at the bottom of the screen. You have the option of viewing the contents of the file before you enter a single number or sequence of numbers by typing "LIST" at either prompt. If you wish to view only a selected portion of the file beginning with the first CAS number, enter "LIST" in the first prompt and then enter the number corresponding to the last CAS number that you want to see at the second parameter.

If you do not want to start at the first CAS number in the file, enter "LIST" at the second prompt, while entering the number where you would like to begin the listing at the first prompt. When you are ready to proceed, enter the starting and ending CAS numbers of the chemicals for which you want to retrieve SMILES notations.

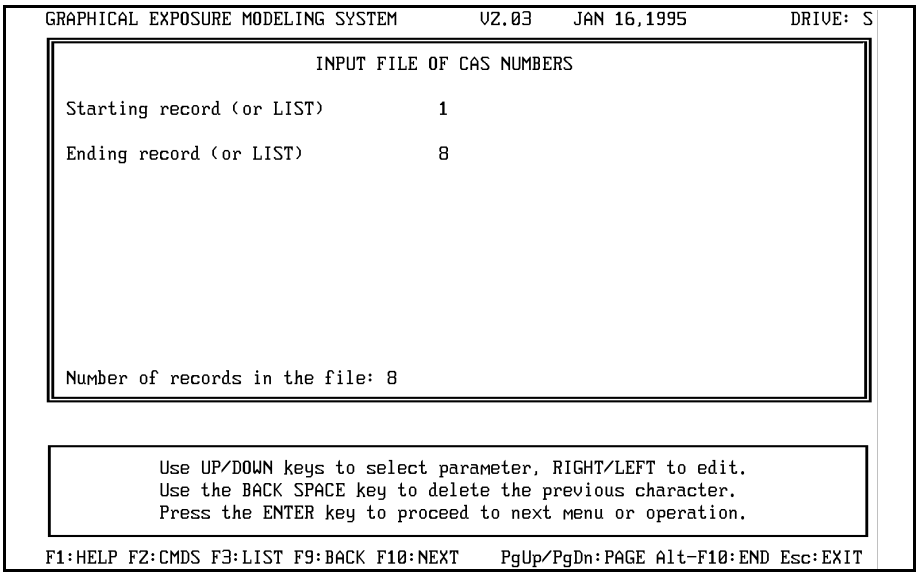

Figure 2-6. Entry of CAS Number from File Menu

Enter the F10 function key when you are done to return to the menu shown in Figure 2-3 for SMIGET. You may select either of the first two options again if you wish to enter additional CAS numbers either directly or from a different file. If you are through entering CAS numbers, select the third option by positioning your cursor over the option and pressing the ENTER key.

The next menu requires you to enter a file label if you chose to save the output. This menu is shown in Figure 2-7 below.

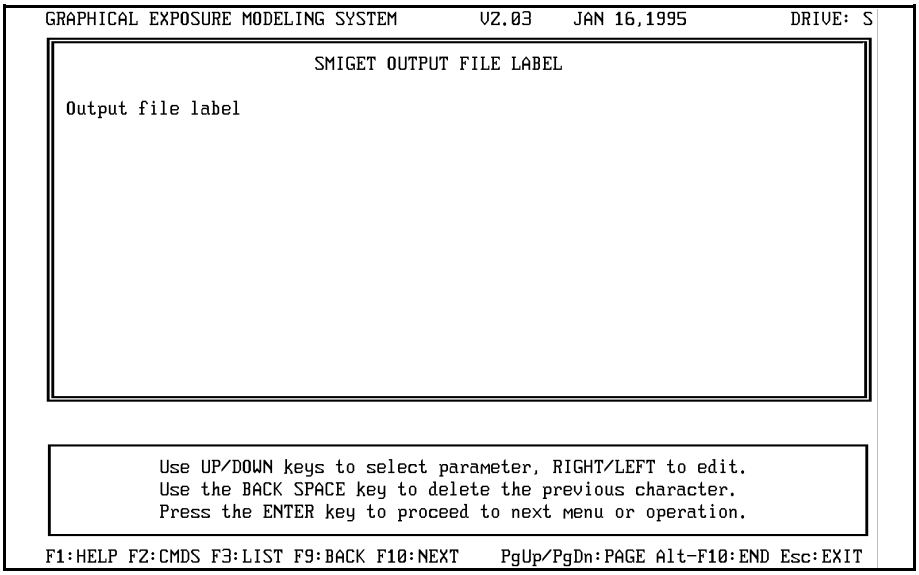

Figure 2.7. SMIGET Output File Label Menu

Enter a file label. It will help you identify the file at a later date in the catalog manager. PCGEMS next checks the dataset path in the PCGEMS configuration file and then searches the directories listed there for the CAS Number/SMILES dataset. If it cannot find the dataset, it will prompt you to enter the file path to the appropriate dataset. (To change the data path in this file, proceed to Subsection 7.2.)

If PCGEMS locates the CAS numbers that you enter in the database, it will show you the SMILES notations, the chemical names and CAS numbers in an output screen, if you chose to view the output. If you chose to save the output, two files are saved: SMILE*.DAT and SMILE*.OUT. The SMILE*.DAT file can be used as input for any of the other chemical property estimation programs.

If PCGEMS cannot find any of the CAS numbers in the CAS Number/SMILES dataset, the following message appears (XXX below is replaced by a sequential number that is assigned according to the number of error files in your directory).

```
Please check CASXXX.ERR in ____.CMF (your default
catalog file name) for CAS numbers error.
```

This file lists all the CAS numbers whose SMILES notations and chemical names

could not be located in the CAS Numbers/SMILES dataset . To see the contents of the file, use the output function in the catalog manager program (Subsection 4.7).

2.3 Retrieve Properties from the Title III Database (TITLE III)

This program provides access to the TITLE III dataset, which contains the already measured or estimated values for 252 of the chemicals listed in Section 313 of the Emergency Planning and Community Right-To-Know Act (Title III of SARA). If the chemical that you enter is one of the 252 chemicals in the TITLE III dataset, an output file of retrieved property values will be created which can be used as input to the modeling programs. If it is not found in the TITLE III dataset, an input file of the CAS numbers is saved for entry into either the PCLOGP or SMIGET programs.

To use this program, you must know the Chemical Abstracts Service (CAS) numbers of the chemicals. CAS numbers may be obtained from the CAS index for chemicals which is available at most libraries. In order to use this program, you must have entered the file path to the TITLE III dataset in the PCGEMS configuration file. When you select this option, the first screen provides you with three options that are provided below.

Option: 1. Display Output to Screen

You may choose to view the retrieved SMILES notations or chemical property values on your PC screen and not save them.

Option: 2. Save Output in a File

You may choose to save and not view the output on the screen. It will be written directly to a file that may be reviewed using the catalog manager.

Option: 3. Save and Display Output

You may choose to both display the model results on screen and save it.

When you select an option, the next menu that appears allows you to enter the CAS

number directly or enter a list of CAS numbers from a file. For the file input option, your file must be in the format provided below:

FILE FORMAT FOR INPUT FILE OF CAS NUMBERS

1. The file must be in ASCII format.
2. The variable containing the CAS numbers must be the first variable in the file (columns 1-13).
3. The file extension must be "DAT".

Both the Direct Input of CAS Numbers and the Input of CAS numbers from a file may be selected in the same session. Once you have entered all the CAS numbers either through the direct CAS entry or through an input file of CAS number or both, select option 3: End input - run.

The CAS numbers that you entered will be checked against the Title III database. All items that are matched will be shown on the screen and/or saved in a file depending on the choice that you made on the first screen. If there are items that do not match from the list of chemicals that you entered, then you will be informed of this in a message like the one shown below:

```
CAS not found in TIII Sec313 database.  Output to
CASXXX.DAT.  This file can be used as input to
SMIGET and PCLOGP.
```

You can use the CASXXX.DAT file mentioned above as input to estimate the properties in PCLOGP and PCCHEM. The CAS numbers that were found in the dataset will be saved in two output files if you chose to save the output, with the same file names as that generated by the PCCHEM program: CHEM*.OUT which is the report file showing the measured properties in tables; and CHEM*.DAT which is a data file that is accessed by the modeling programs.

2.4 Octanol/Water Coefficient (PCLOGP)

The octanol/water partition coefficient is the equilibrium concentration of solute in a non-polar solvent (octanol) divided by the concentration of the same species in a polar solvent (water). It will be referred to in this manual with the abbreviation K_{ow} and is often expressed as a logarithmic value (log K_{ow}). Log K_{ow} is a measure of a chemical's tendency to partition itself between an organic phase and an aqueous phase. This measurement gives you an idea of the chemical's tendency to partition itself between an organic phase such as fish or soil and an aqueous phase such as the stream or pore water. High K_{ow} values indicate a hydrophobic chemical, while low values of K_{ow} indicate a hydrophilic chemical. Measured log K_{ow} values, although preferable to calculated values, are time-consuming, expensive and sometimes difficult to obtain. This program provides a reliable means of estimating this property.

The menu shown in Figure 2-8 appears when you choose this program.

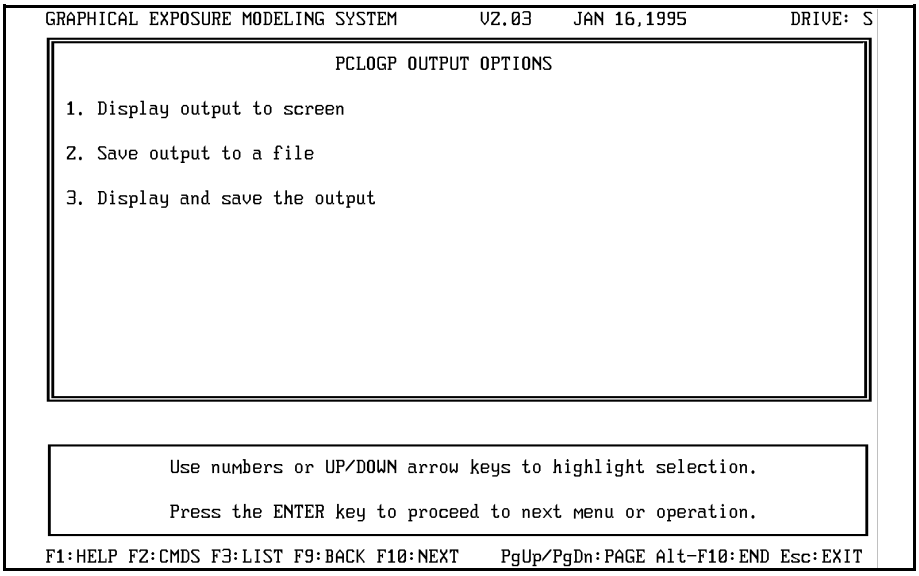

Figure 2-8. PCLOGP Output Options Menu

Choose an output option by positioning your cursor over the option. If you choose the second option, the output will not be displayed on the screen; instead, it is written directly to a file which you may view using the catalog manager.

If you choose to save the output, the menu shown in Figure 2-9 appears (otherwise, the program proceeds to the menu shown in Figure 2- 10). Enter a descriptive label which will allow you to identify the file contents later. When you are done, press the ENTER key.

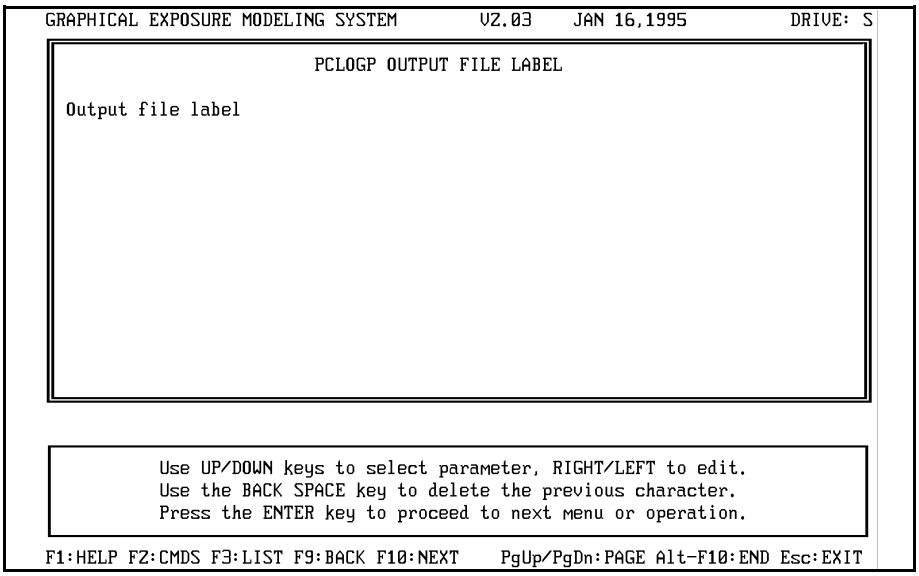

Figure 2-9. PCLOGP Output File Label Menu

The navigational menu shown in Figure 2-10 follows.

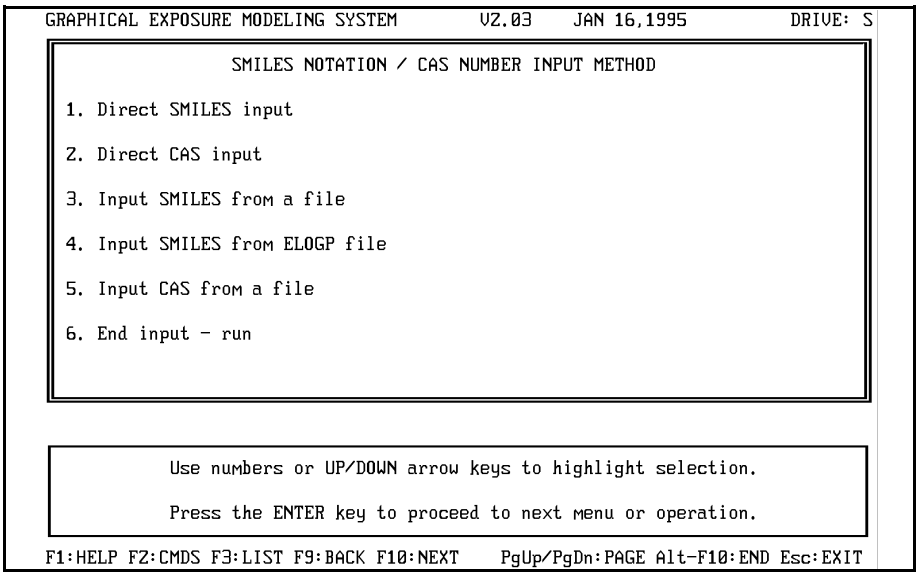

Figure 2-10. PCCHEM Input Method Options Menu

This menu is the main navigational menu in PCLOGP. The six options available to you are listed below, together with the menus that appear when you select each one.

Option 1. Direct SMILES Input:

This option allows you to enter the SMILES notation of a compound directly. When you choose this option, the menu in Figure 2-11 appears. Enter the SMILES notation in the first field on the screen. The CAS number and chemical name are optional and should be entered, if known.

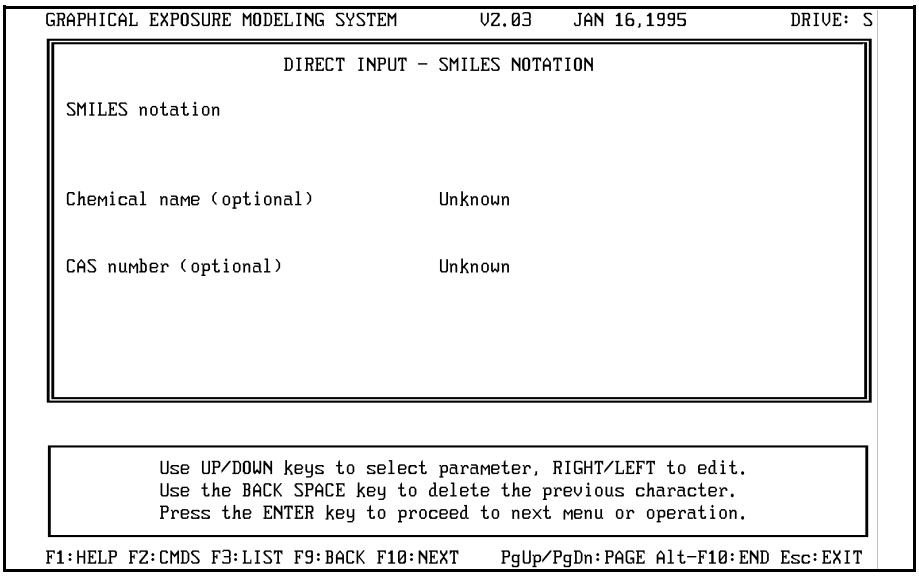

Figure 2-11. Direct SMILES Notation Entry Menu

SMILES is an acronym that stands for Simplified Molecular Input Line Entry System. It is a linear, chemical structure notation system that is the preferred way of describing a chemical structure for both GEMS and PCGEMS. It has become the preferred language for chemical nomenclature because it is both easy to learn and flexible. A non-chemist can learn the nomenclature as easily as a chemist can. For many chemicals, there are many ways to describe the structure, each way being equally correct. SMILES uniquely describes a chemical structure with strings of characters, using a small vocabulary of symbols. There are a few rules to remember when writing SMILES:

- Atoms are represented by their atomic symbols. The first letter of the atomic symbol must be in uppercase, except in the case of aromatic carbon atoms. The SMILES interpreter automatically determines the number of hydrogens attached to each atom as long as the hydrogen atom has a neutral charge, so you don't have to indicate them. Therefore, the presence of hydrogen is omitted from the notation. The valid atomic symbols are Br, C, Cl, F, I, N, O, P and S.

- In SMILES, all bonds are assumed to be covalent unless otherwise indicated.

Bonds are represented by characters which separate atomic symbols. The bond character for single bonds is the en dash (-) (these do not have to be indicated since they will be assumed whenever two atoms are adjacent). For double bonds use the "equals"(=) sign, for triple bonds use the pound (#) sign, and for delocalized bonds use the period (.) sign. For aromatic bonds, the bond characters are alternating single bond (-) and double bond (=) symbols. Alternatively, you can use the lower case letter "c" to specify aromatic carbon atoms. If adjacent atoms are both aromatic, an aromatic bond is assumed. Attached hydrogens and formal charges are always specified inside brackets.

- A branch off the main molecular structure is indicated by enclosing the branch atoms in parentheses (). Nesting is allowed up to ten levels.

- To indicate a cyclic structure in SMILES, break one single or aromatic bond in each ring. Assign a number to that bond, and label the carbons on either side of the "broken" bond with that number. The SMILES notation for benzene, a cyclic structure, is thus c1ccccc1.

- The SMILES notation for the molecule ends at the occurrence of the first space.

Some examples of the SMILES are:

CCCO	propane	C3H7OH
c1ccccc1O	phenol	C6H5OH
C1CCCCC1O	cyclohexanol	C6H11OH
Cc1ccc(C)cc1	p-xylene	C8H10

For more information on SMILES, refer to the MedChem Software Manual, Release 3.32, from the Medicinal Chemistry Project of Pomona College in Claremont, CA [2].

When you have entered the information required by this menu, press the ENTER key. You will be returned to the menu shown in Figure 2-10, where you may select another input method option to enter an additional SMILES notation or you may run the program. Note: You may enter as many structures as you choose before running the program.

Option 2. Direct CAS Input

This second input option allows you to identify the chemical by directly entering the CAS number of the chemical (Figure 2-12).

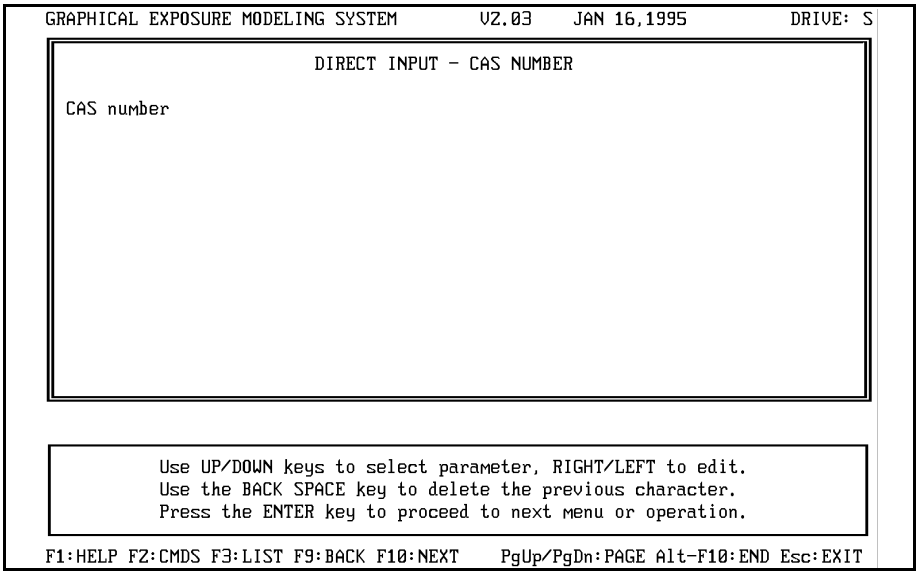

Figure 2-12. Direct CAS Number Entry Menu

The CAS number may be entered with or without blanks or hyphens. When you enter a CAS number, it is checked against the CAS Number/SMILES dataset. If the number is not found there, you are advised of this fact. Next, you will be returned to the menu shown in Figure 2-10, where you may enter additional structures using different entry methods if you choose, or you may run the program.

NOTE: You must have access to the CAS Numbers/SMILES dataset if you plan to use this option since the CAS number will be cross-referenced to this dataset to obtain the necessary SMILES notation and chemical name (if the CAS number is found in the dataset). An additional menu will appear, asking for the path to the CAS Number/SMILES dataset if the dataset is not located in the PCGEMS configuration file. This menu is not illustrated. If you wish to change the PCGEMS configuration file to include the CAS Number/SMILES dataset, proceed to Subsection 7.2. The file path that you enter here tells PCGEMS how to proceed to where the file is

located. For example, if the particular datafile in the CAS Number/SMILES dataset that is asked for (e.g., SMILES1.BIN) is in drive A:, you should enter "A:" telling PCGEMS that the file that is asked for is located in drive A:.

- *Option 3. Input SMILES From a File*

  This option allows users to specify a file containing SMILES notation for use as input for the program. The input SMILES file may be created either by running the SMIGET program as an input file or through user-entry. If you are using a file that you have created yourself, the file must be in ASCII format and it should have the file format provided below (The SMILES notation must be the first variable):

VARIABLE	VARIABLE TYPE
SMILES Notations	Alphanumeric (Maximum 120 characters)
Blank Space	
Chemical Name	Alphanumeric (Maximum 124 characters)
Keyword "CASNO"	5 characters
Blank Space	
CAS Numbers	Numeric (Maximum 13 characters)

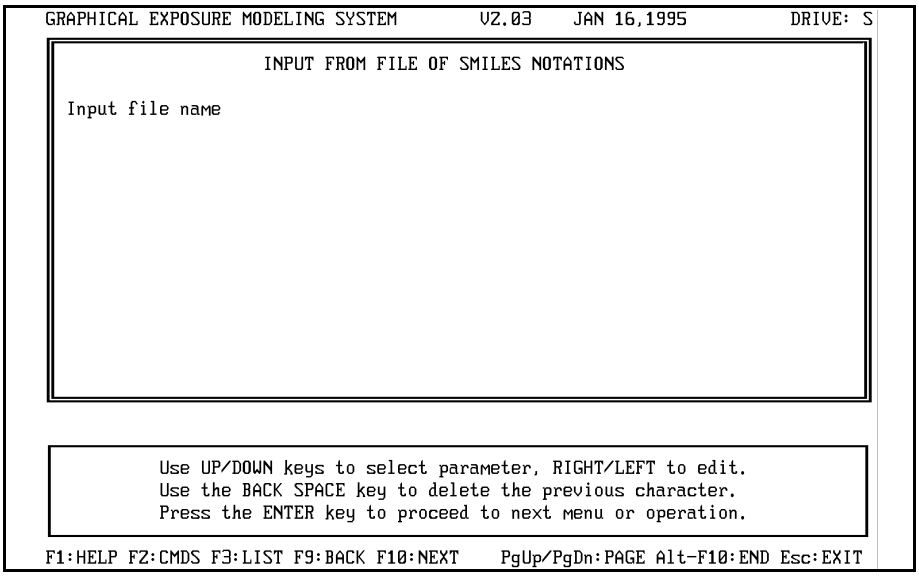

Figure 2-13. SMILES Notations File Entry Menu

Enter the name of the files containing the SMILES notations. Do not enter the file extension since it is assumed that the file extension is ".DAT". PCGEMS will read the input file and automatically display the starting and ending structure numbers in the file. This information will appear in a menu following the menu in Figure 2-13 and is similar to the menu shown in Figure 2-7. You may choose to enter a single structure from this file by entering the same starting and ending numbers or you may enter a range of numbers by entering the starting and ending structure number in the range that you wish to estimate. If you wish to see the contents of the file, enter "LIST". You may list only a certain portion of the file by following the same format as you would in entering a range of structures except that one of the entries, either the starting or ending number, should be "LIST". (This menu is not illustrated since, except for the title, it is exactly like the menu in Figure 2-7).

When you are through entering the structure number(s), press the ENTER key. You will then be returned to the main PCLOGP navigational menu. You may select another input option, or you may choose the "End of Entry" option to begin the run.

- *Option 4: Input SMILES from ELOGP File*

This option allows you to bring SMILES structures from the output file created by the ELOGP program. This $logK_{ow}$ estimation technique was developed by Syracuse Research Corporation. The ELOGP output file must be in the same format as the file that was created by the PCSMIGET program. The file format is provided below:

FILE FORMAT

The input file must be in ASCII format. The SMILES notations must be in the first column of the file. There must be three columns that are separated by a blank space. The format of this file is provided below:

VARIABLE	VARIABLE TYPE
SMILES Notations	Alphanumeric (Maximum 120 characters)
Blank Space	
Chemical Name	Alphanumeric (Maximum 124 characters)
Keyword "CASNO"	5 characters
Blank Space	
CAS Numbers	Numeric (Maximum 13 characters)

An example of what your file should look like is provided below:

```
c(cccc1)(c1)-CC Benzene, ethyl- CASNO 100414
c(cccc1)(c1)-C Benzene, methyl- CASNO 108883
C(=CCl)(Cl)Cl Ethene, trichloro- CASNO 79016
```

When you select this option, the menu shown in Figure 2-14 appears.

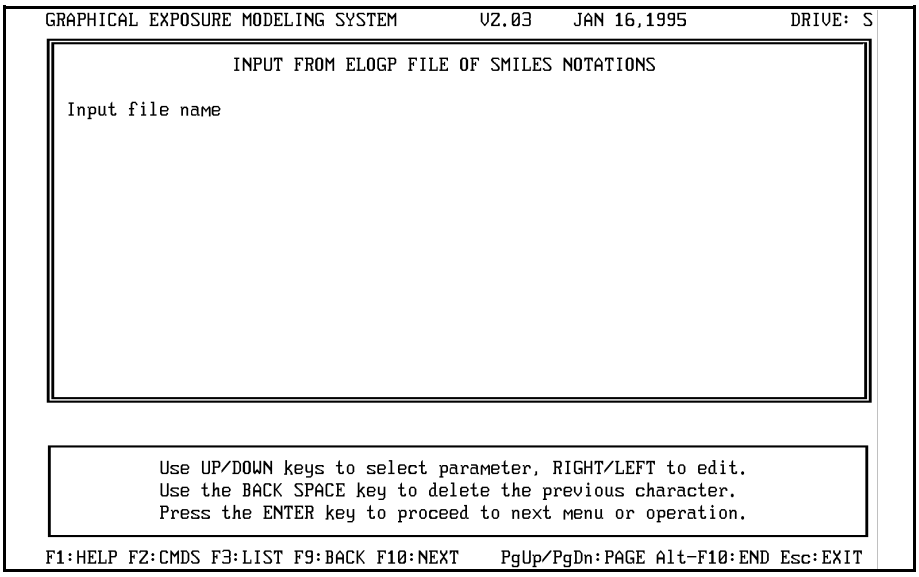

Figure 2-14. ELOGP Input File Menu

- *Option 5. Input CAS From a File*

This option allows users to specify a file containing CAS numbers for use as input to the program. When you choose this option, the menu shown in Figure 2-15 appears. Enter the name of the input file of CAS numbers without its extension. If you are entering your own file of CAS numbers, the file must have the format provided below:

FILE FORMAT

1. The file must be in ASCII format.
2. The file extension must be .DAT.
3. The variable containing the CAS number must be the first variable in the file and it can only be 13 characters long.

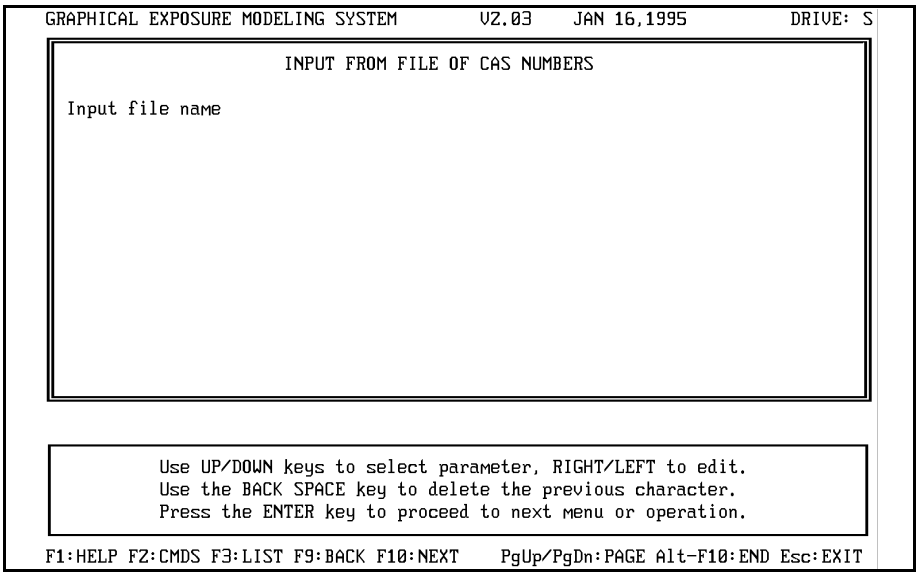

Figure 2-15 CAS Numbers File Name Menu

When you have entered a file and after PCGEMS has verified that it is in the correct format, PCGEMS will display an additional menu containing the starting number and ending numbers of the range of CAS numbers in the file. The menu is in the same format as the one in Figure 2-7 and will not be illustrated here. You may enter a single CAS number by entering the same number for the beginning and ending structure number prompt. Alternatively, you may enter a range of numbers by entering the beginning and ending structure numbers in the range that you want to estimate. You may enter "LIST" at either prompt to see the contents of the file or a portion thereof.

Note: You need the CAS Number/SMILES dataset if you select this option. If this dataset is not located in the data path in the PCGEMS configuration file that you created, an additional menu will appear, asking you to enter the path to one or more of the six data files that make up the SMILES dataset (depending on the range of structures that you enter). This menu will not be illustrated.

The numbers that are in the file are checked against the CAS Number/SMILES dataset. If any of the CAS numbers are not found in the dataset, you will be informed of this fact and shown the numbers that could not be found. These

chemicals will not be included in the PCLOGP run unless you input their structure using SMILES notation. You will be returned to the main menu, where you may select other input options or run the program.

- *Option 6. End Input - Run*

  When you are through entering chemicals by SMILES notations or by CAS numbers, enter this option to begin the PCLOGP run. You will see a message in the instruction section of your menu advising you that a batch file is being created, followed by the program herald. When the processing is complete, you will see the output results. If you elect to save the output in a file, you will be shown the output file name in the instruction section. The output file name is automatically generated by PCLOGP and is named CLOGPXXX.DAT (where XXX is replaced by a three-digit number that is sequentially assigned according to the number of datasets that have been created in that directory). You may use this file as an input file for the PCCHEM program. The output file is then inserted into your active catalog file. You will also be advised to check a file called the PCLOGP.ERR file if there are some structures for which the log K_{ow} values could not be estimated. Use the output function in the catalog manager program (Subsection 4.7) to view this file.

2.5 Chemical Property Estimation (PCCHEM)

The Chemical Property Estimation Program facilitates estimation of chemical properties without requiring any knowledge about organic chemistry. The only input necessary to run PCCHEM is the SMILES notation for each chemical, and its log K_{ow} value. The log K_{ow} value is only necessary if you are planning to estimate the water solubility, the organic carbon partition coefficient, and the bioconcentration factor. If you do not know the log K_{ow} value for the chemical, use the PCLOGP program to calculate it before using this program. You can use a PCLOGP output file as a source of both the log K_{ow} value and the chemical structure for the chemical for use in PCCHEM. PCCHEM provides estimation routines for the following properties: boiling point, melting point, water solubility, vapor pressure, Henry's Law Constant, organic carbon partition coefficient, and bioconcentration factor. These properties are required information for a variety of environmental models. The PCCHEM program will scan a molecular structure that you enter, determine which

fragments are present and in what quantity, and select the most reliable equation to apply from the equations available in the CHEMEST program.

PCCHEM has certain limitations as to the chemicals whose properties it can estimate. PCCHEM cannot estimate properties for polymers, salts, inorganic compounds, or mixtures. Furthermore, values for some molecular fragments are as yet unavailable. When a molecule contains such a fragment (e.g., the isocyanate group), all estimations which depend on fragment values (octanol/water partition coefficient, water solubility, Henry's Law Constant, organic carbon partition coefficient, and bioconcentration factor) must be considered incomplete.

NOTE: PCCHEM will automatically scan a chemical when it is entered by the user and check it against the GEMS CHEMEST dataset of measured properties for 856 different chemicals. If the chemical that is entered by the user is found here, then all the measured properties for the chemical that are in the dataset are retrieved and the remaining properties that are requested by the user are estimated.

The methods used in PCCHEM to estimate values for each property and a brief note on their applicability and accuracy are given below. The information about the methods is contained in the Handbook of Chemical Property Estimation Methods [3]. Refer to it if you want more detailed information about the estimation methods. Additional information is available in the Automated Procedures for Physiochemical Property Estimation User's Manual [4].

Melting Point

Two methods are included in PCCHEM to estimate the melting point of a chemical, which is the temperature at which a solid substance melts or fuses. These methods are the Lorenz and Herz Method, and the Grain and Lyman Method. The former estimates the melting point using as input the normal boiling point, while the latter depends on the liquid density and a constant related to chemical class and molecular weight. The Grain and Lyman method was included because it is more reliable and accurate in the estimation of the melting point of chemicals which include atoms other than C, H, N, O, S, F, Cl, Br, or I. For all other chemicals, the Lorenz and Herz method is used.

Boiling Point

Boiling point is defined as the temperature at which the vapor pressure of a liquid is equal to the pressure of the atmosphere on the liquid. Besides being an indicator of the physical state of a chemical (liquid vs. gas), the boiling point also provides an indication of the chemical's volatility. Two methods to estimate boiling point have been implemented in PCCHEM: the Meissner Method and the Miller Method. The Meissner Method is applicable to a wide variety of chemicals and is quite accurate, having an average error of 2% and a maximum error of 8% in degrees Kelvin (K) at boiling point. However, the Meissner Method cannot estimate the boiling point for chemicals with the nitro group since molar refraction information for the nitro group is not available. The Miller method, while less accurate than the Meissner Method, is able to estimate the boiling point for chemicals with the nitro group.

Water Solubility

Water solubility is the maximum amount of a chemical that will dissolve in pure water at a specific temperature. Water solubility is one of the most useful chemical parameters for liquids and solids, since it can give much insight into the fate and transport of an organic chemical in the environment. The more soluble a chemical is, the more rapidly and extensively it will be distributed through water in rivers, lakes or in saturated soil. Less soluble chemicals tend to be either absorbed by soils and sediments or accumulated within fish and microorganisms. Chemicals that are highly soluble are also more likely to be biodegraded by microorganisms in soils, surface water and sewage treatment processes.

PCCHEM uses seven different equations which correlate water solubility with the octanol/water partition coefficient (log K_{ow}). These equations are applicable to a wide variety of chemicals. PCCHEM will most often estimate water solubility using one of three of these equations since they apply to most chemicals, including acids, and because a recent study has shown that these three equations provide the best overall estimates. These equations are shown below. (Note: The third equation is applicable to acids only and is further limited to acids whose log K_{ow} values fall between -0.5 to 3.2.)

$$S = 10(-1.1123 \log K_{ow} + 0.686 - 0.0099(T_m-25))$$
$$S = 10(-1.034 \log K_{ow} + 0.455 - 0.0099(T_m-25))$$
$$S = 10(-0.65 \log K_{ow} + 0.279 - 0.0099(T_m - 25))$$

where

K_{ow} = octanol/water partition coefficient

T_m = melting point (K)

Vapor Pressure

Vapor pressure is defined as the pressure of a vapor, at any given temperature, in equilibrium with its liquid or solid form. When a chemical has been spilled, knowing its approximate vapor pressure helps to estimate its rate of evaporation. Two methods have been included in PCCHEM to estimate vapor pressure. These methods are the Antoine Method and the Modified Watson Method. The Antoine Method is restricted to liquids and gases and is generally applicable over the pressure range from 760 mm to 10^{-3}mm. The Modified Watson Method is restricted to liquids and solids and is applicable from 760 mm to at least 10^{-7}mm. Both of the methods require normal boiling point and K_f, a structure-dependent constant that is derived by PCCHEM from a consideration of the dipole moments of polar and nonpolar compounds.

Henry's Law Constant

Henry's Law Constant relates the equilibrium concentration of a compound in the gas phase to its concentration in the liquid phase as a direct ratio. This property is the key factor in determining a chemical's degree of volatilization from water. Knowledge of volatilization rates is necessary to determine the amount of the chemical that enters the atmosphere and, as a result, the change of pollutant concentrations in water bodies. The larger the value of this property, the more rapidly a chemical is likely to volatilize from water. The method used in PCCHEM uses the vapor pressure and the water solubility ratio to estimate Henry's Law Constant. Henry's Law Constant estimations are estimated at 25°C. If the chemical is a gas, vapor pressure is assumed to be 760 mm of Hg.

Bioconcentration Factor

The bioconcentration factor (BCF) of a chemical indicates the degree to which a chemical dissolved or suspended in water may accumulate in aquatic organisms, usually in fish. Values for this property are usually expressed as the ratio of the concentration in the organism (wet weight) to the concentration of the chemical in water.

In estimating the bioconcentration factor, the following regression equation is used

in PCGEMS:

$$\log BCF = 0.76 \log K_{ow} - 0.23$$

where, K_{ow} is the octanol/water partition coefficient.

This equation is the only one available in PCCHEM. The Handbook of Chemical Property Estimation Methods provides a comparison of estimated values with laboratory measurements of bioconcentration factor which indicate that BCF's estimated using K_{ow} fall within an order of magnitude of the measured value. Also, K_{ow} estimation has been studied more extensively than other techniques.

Organic Carbon Adsorption Coefficient

Adsorption of a chemical to soils and sediments can be described as a partitioning of the chemical between the aqueous phase (the water in which the chemical is dissolved) and the solid phase (the soil or sediment into which the chemical solution comes in contact). This partitioning is described using the organic carbon partition coefficient (K_{oc}), which can indicate the extent to which an organic chemical partitions itself between solid and solution phases. The following regression equation is used to estimate Koc:

$$K_{oc} = 0.544 \log K_{ow} + 1.377$$

where, K_{ow} is the octanol/water partition coefficient

The menu shown in Figure 2-16 appears when you select PCCHEM:

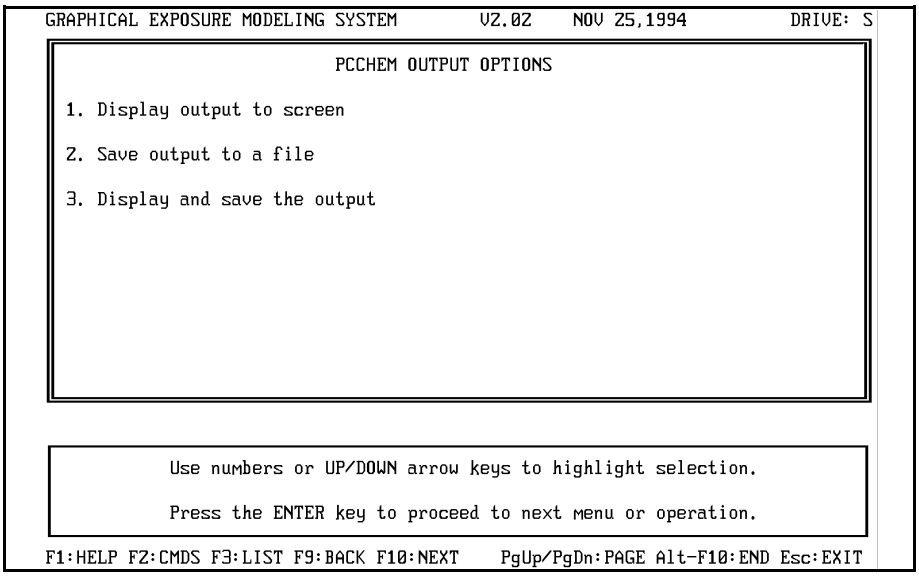

Figure 2-16. PCCHEM Output Options Menu

You have three output options. PCCHEM is different from the other estimation programs in PCGEMS in that these options may be changed several times during a session. Each time you change an output mode in PCGEMS, it means that all the structures entered after the change will adhere to the new output mode, while all structures entered prior to the change will follow the previous output mode. This is important since in PCGEMS, output mode 2, "Save output to a file" is the only mode which allows you to enter a range of structures from a file at one menu. For the other output modes, if you wish to estimate a range of structures from a file, you must enter all the structures in the range one structure at a time. For more detail on methods, refer to the Automated Procedures for Physiochemical Property Estimation User's Manual.

When you have selected an output option, you will see the menu shown in Figure 2-17. This menu is shown as it would appear if you chose output option 3. If you had chosen option 1, only the first two prompts would appear. Similarly, if you had chosen option 2, only the third prompt would appear.

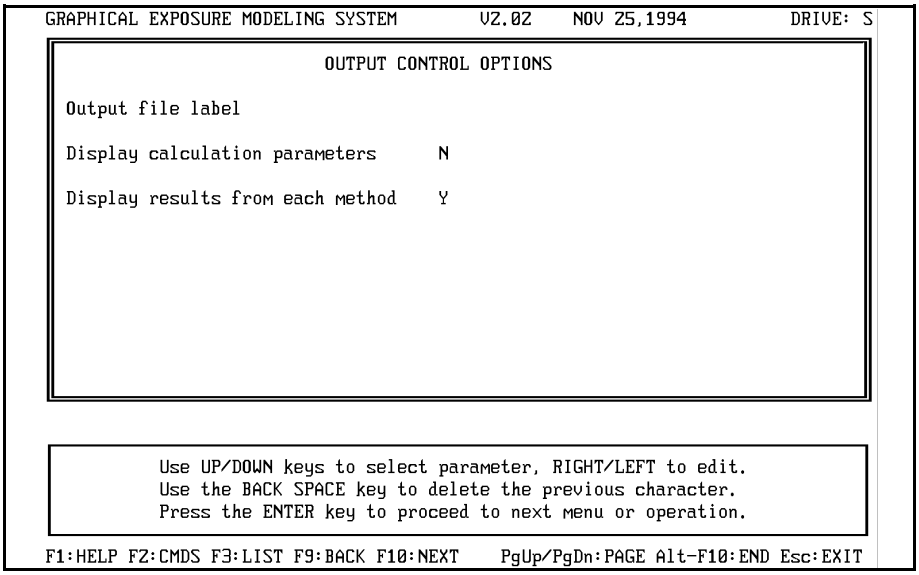

Figure 2-17. Output Control Options Menu

Three properties; melting point, boiling point, and vapor pressure, also depend on calculations based on molecular weight, liquid density and other structural elements as part of their estimation methods. You may choose to see these parameters by changing the default value for the first option from “N” to “Y”. As explained previously, the properties in PCCHEM are each estimated using several different methods. You may review the results that are generated by each method before the most accurate method for that chemical is automatically selected by PCGEMS in a summary table by leaving the second option, “Display results from each method”, at the default of “Y”. The third option on the screen is the output file label. Enter a label that will help you recognize the contents of this output file later.

When you are through with this menu, press the ENTER key and the menu shown in Figure 2-18 appears. This is the main navigational menu in PCCHEM. Each option is detailed below, along with the menu that will appear when you select each option. You may use either or both of the first two options in any sequence to enter as many chemical structures as you wish.

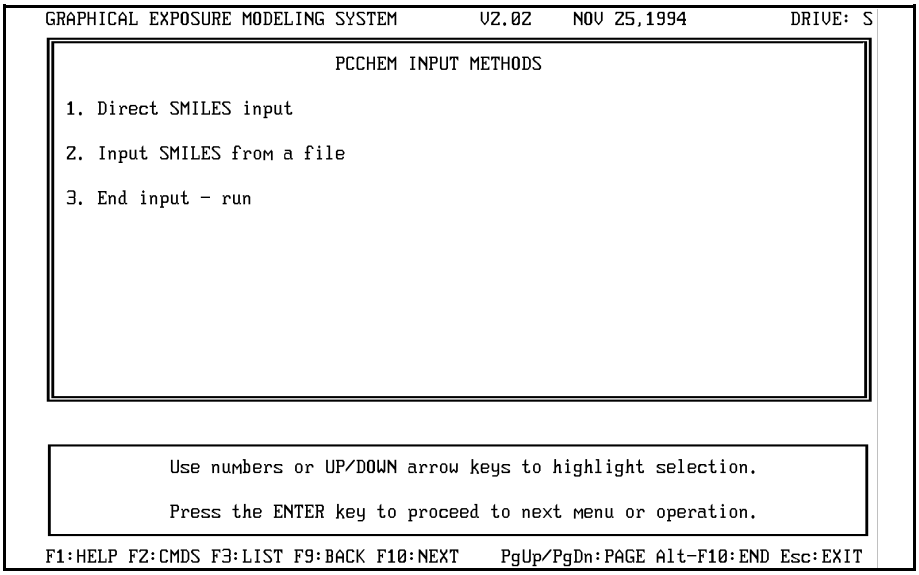

Figure 2-18. PCCHEM Input Methods Menu

Option 1. Direct SMILES Input

When you enter this option, the menu shown below in Figure 2-19 appears. Enter the SMILES notation. SMILES is an easy system to learn. For the rules on using SMILES, refer to Subsection 2.3 of this chapter. Enter the chemical name and CAS number, if you know it. Chemical names and CAS numbers are not required for the estimation routines. However, they are useful if you are planning to save the results, since they will help identify the chemical structure. If you are unfamiliar with SMILES but know the CAS number, you may use the SMIGET program to obtain the correct SMILES notation. Refer to Subsection 2.2 for information on this program.

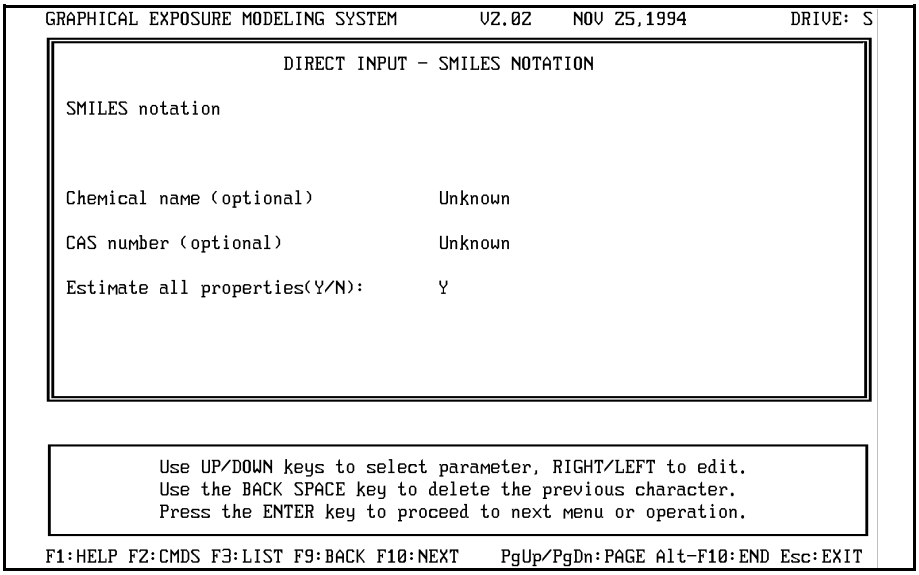

Figure 2-19. Direct Input - Smiles Notation Menu

Accept the default Y if you want PCCHEM to estimate all properties. The menu shown in Figure 2-20 appears next. If you have measured values for some properties and do not wish to estimate all the properties, change the default response to N for NO in the menu shown in Figure 2-19. The default values in the menu shown in Figure 2-20 would then all be N.

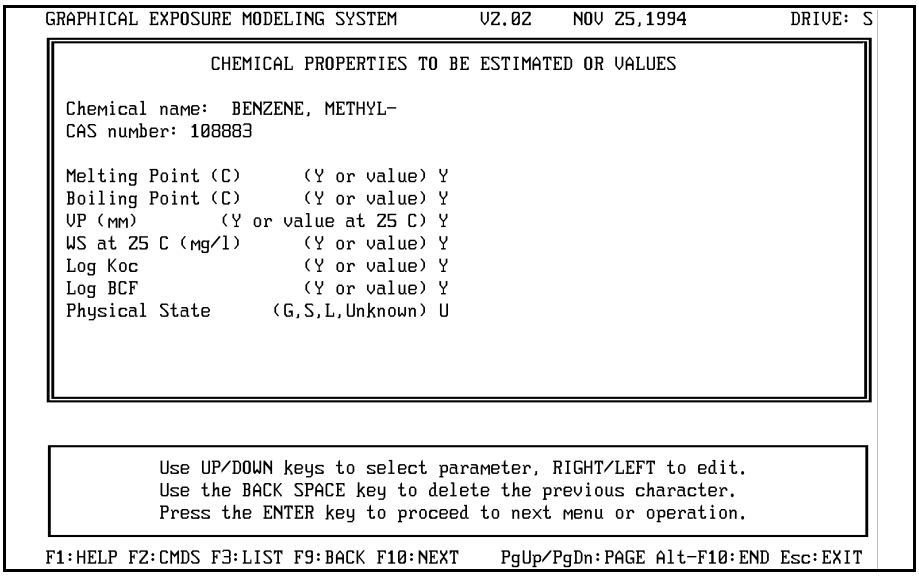

Figure 2-20. Property Estimator Indicator Menu

At this menu you have the opportunity to provide measured values for the properties shown on the screen. Some of the properties on the screen are estimated using other property values as input. For example, the Henry's Law Constant is estimated using water solubility and vapor pressure. Providing measured values for these properties will increase the reliability of the resulting estimate. The measured values must be provided in the units indicated in parentheses beside each property. (Values for bioconcentration factor are unitless.) If you have measured values in different units, use the conversion table provided in Table 2-2 to obtain values in the correct units. If you know the physical state (solid, liquid, or gas) of the chemical of interest, it should also be provided at this menu. Otherwise, PCCHEM will assign a physical state based on the estimated melting and boiling points.

Table 2-2. Conversion of Units

Melting Point and Boiling Point		
From:	KELVIN (K)	FAHRENHEIT (F)
To: CELSIUS (C)	(K - 273.16)	(C x 1.8) + 32
Vapor Pressure		
From:	ATM	KPA
To: mm	ATM x 760	KPA x 7.75074
Water Solubility		
From:	Mol/L (MOL)	
To: mg/L	MOL x MW* x 1000	

* MW is the molecular weight of the chemical.

Depending on your entries to this menu, several different menus will appear when you press enter to continue. For many of these properties, PCCHEM requires a

measured or estimated log K_{ow} value. You enter this value in the menu shown in Figure 2-21 shown below. If you are estimating vapor pressure, but not Henry's Law Constant, an additional prompt will appear (as illustrated in Figure 2-21) at which you can select the temperature at which you want the vapor pressure to be estimated. (If you are also estimating Henry's Law Constant, the temperature will be assumed to be 25 degrees Celsius.)

The menu shown in Figure 2-21 appears next.

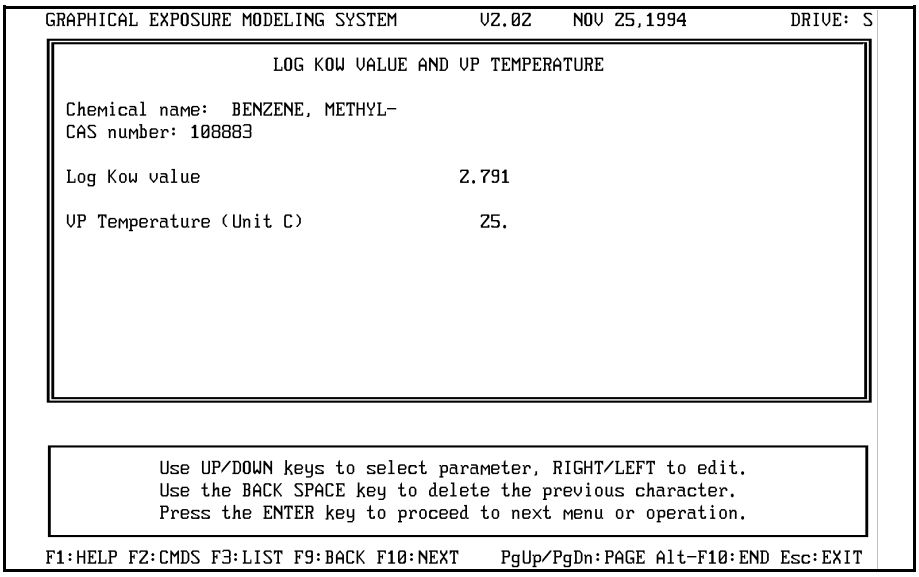

Figure 2-21. Log K_{ow} Value Menu

Most users of PCCHEM do not have measured values for log K_{ow} available, and use the PCLOGP program to estimate it. If you do not know the log K_{ow} value, exit from this program and use the PCLOGP program to estimate the log K_{ow} value. Note that the output file created by PCLOGP program contains the log K_{ow} value and may be entered as an input file for this program. The method for doing this is explained in option 2, described below. This menu also asks you for the Vapor Pressure temperature. If you wish to estimate a chemical's vapor pressure at anything other than the standard 25° C, enter the temperature you would like here. The chemical name and CAS number will appear in this screen, if the program has this information.

When you have entered a log K_{ow} value, you will be returned to the main PCCHEM navigational menu shown in Figure 2-18. At this point you may enter another SMILES notation directly or enter a range of structures from a file, or run PCCHEM.

Option 2. Input SMILES From a File

The menu in Figure 2-22 appears when you choose this option. At the first prompt, enter the name of the file containing the log K_{ow} values. If the input file you wish to use was created by PCLOGP, the file name will have the form CLOGPXXX.DAT. You do not need to include the .DAT extension here.

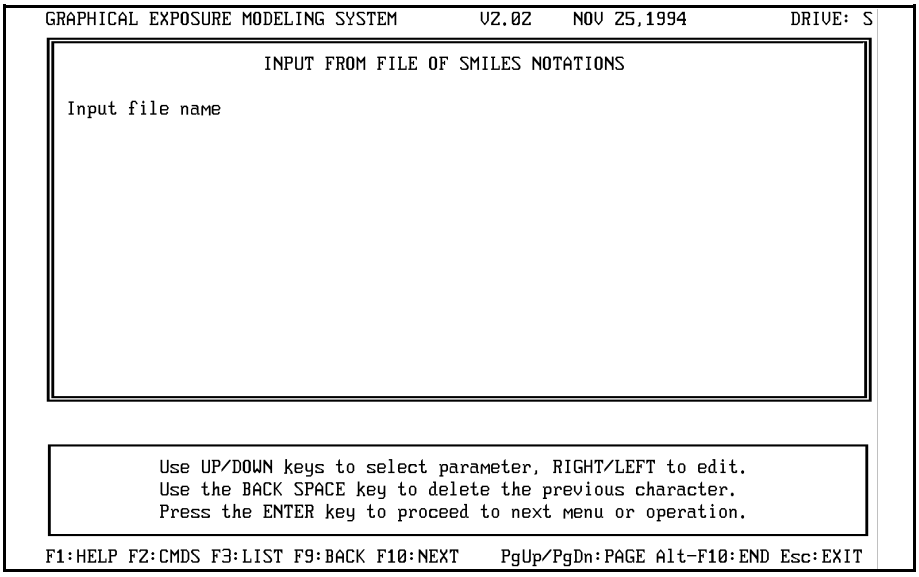

Figure 2-22. SMILES Notation File Entry Menu

Depending on the output option you selected from the Output Options menu at the beginning of a PCCHEM run (see Figure 2-16), a series of menus will appear prompting you for information the program needs before it can run. Each menu series is profiled in the following paragraphs.

Display Output to Screen

If you chose this option from the Output Options menu, PCCHEM results will be displayed on your PC screen, but they will not be saved. In this case, the first menu that appears in response to selecting "Input SMILES from a File" is similar to the one shown in Figure 2-23; the difference is that the prompt for selecting a range of

structures is not included. Once you have identified the input file, PCCHEM moves to a menu titled "Input from File of SMILES Notations." This menu is illustrated in Figure 2-7, which appeared earlier in this chapter. At this menu, the user enters the number of the structure within the input file for which properties are to be estimated. The structure number is the chemical's numerical position in the file. Use the LIST command at this prompt and you will be provided with a list of the SMILES notations in the file beside the number associated with each.

A YES answer to "Estimate all properties?" will result in all properties being estimated except for any measured values provided by the user. A menu like the one in Figure 2-30 will appear, allowing you to enter measured values (Be sure to read the paragraphs that follow Figure 2-20 for information on providing measured values.) A NO response at the menu shown in Figure 2-20 will limit the estimation to only those properties you identify in the menu shown in Figure 2-20. The final menu in this series is illustrated in Figure 2-21, where you will be required to enter the log K_{ow} value.

Remember, a log K_{ow} is needed to estimate most properties in PCCHEM, but this menu will not appear if you only request the estimation of properties which do not rely on log K_{ow}. If your input file includes a log K_{ow} value, it will automatically appear at this prompt. If you need values for log K_{ow}, it can be estimated using the PCLOGP program. You will have to exit PCCHEM (using the Alt-F10 key) in order to run PCLOGP.

When PCCHEM has all the information it requires, you will be returned to the PCCHEM Input Methods menu shown in Figure 2-18. At this point you may choose to enter information for additional chemicals, or can run the program.

Save Output to a File

If you chose this option at the Output Options menu, the first menu to appear is shown in Figure 2-23. This is the only method with which you are able to generate estimates for a range of structures, since for all other input and output combinations, PCCHEM performs estimates for one chemical at a time. In the first subsection below we discuss the menu series that appears if you choose to estimate properties for a single chemical, and in the second subsection we discuss the menus that appear when you choose to estimate properties for a group of chemicals at once.

1. Single Chemical Estimation

Accept the default response NO to the "Select a range of chemicals?" prompt if you only want properties estimated for a single chemical. PCCHEM will move to the menu illustrated in Figure 2-7, shown earlier in this chapter. At this menu, enter the number of the structure within the input file for which properties are to be estimated. The structure number is the chemical's numerical position in the file. Use the LIST command at this prompt and you will be provided with a list of the SMILES notations in the file beside the number associated with each. Once you have selected the chemical of interest from the file, you will have to indicate if you wish to estimate all the properties.

A YES answer to the "Estimate all properties?" prompt will result in all properties being estimated except for any measured values provided by the user in the next menu. A NO response will limit the estimation to only those properties specified by the user in the next menu. The next menu, shown in Figure 2-20, contains a list of available properties to be estimated. (Be sure to read the paragraphs that follow Figure 2-20 for information on providing measured values.) The final menu in this series is illustrated in Figure 2-21. Remember, a log K_{ow} is needed to estimate most properties in PCCHEM, but this menu will not appear if you only request the estimation of properties which do not rely on log K_{ow}. If your input file includes a log K_{ow} value, it will automatically appear at this prompt. If you need a value for log K_{ow}, it can be estimated using PCLOGP. You will have to exit PCCHEM (using the Alt-F10 key) in order to run PCLOGP. Press return once this menu is completed, and you will be returned to the PCCHEM Input Methods menu shown in Figure 2-18. Here you may enter information for additional chemicals, or start the program.

2. Chemical Group Estimation

If you wish to have PCCHEM estimate properties for a group of chemicals, accept the default, YES to the "Select range of structures?" prompt. We recommend that the input file used for this option include log K_{ow} values, since the user will not have an opportunity to provide this information interactively. (Output files generated by PCLOGP are ideal for this purpose, since they contain both the SMILES notation and estimated log K_{ow} values for each chemical entered by the user.) PCCHEM will proceed to the menu shown in Figure 2-23. This menu will automatically show the numbers for the starting record (usually

1) and for the ending record (usually identical to the number of chemicals in the file). If you wish to see the chemicals included in the input file, type the command LIST at either prompt, and a numbered list of the SMILES notations in the file will appear.

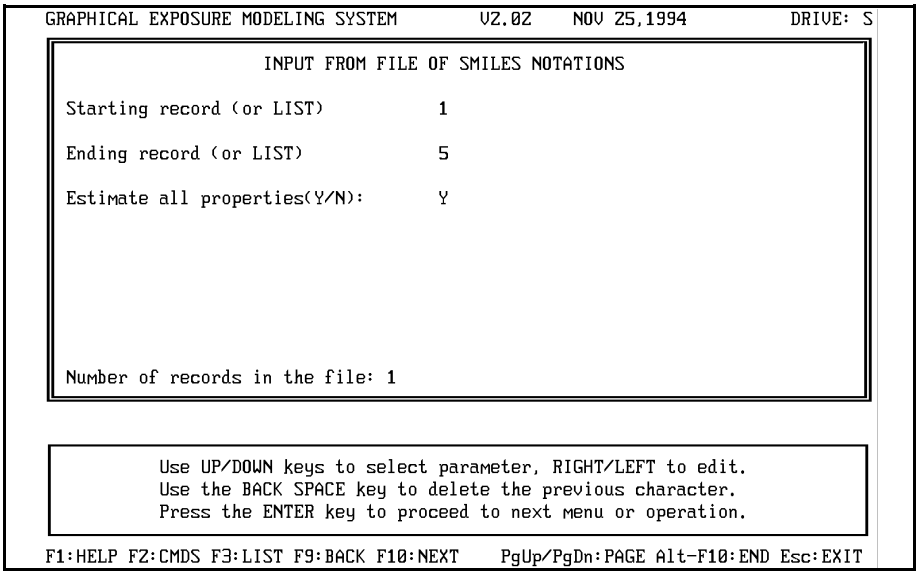

Figure 2-23. Range of SMILES Structures Entry Menu

A YES response to the prompt “Estimate all properties?” completes this menu and returns you to the PCCHEM Input Methods menu illustrated in Figure 2-18. At this point you may run the program or provide information for additional chemicals. If you enter NO in response to this prompt, PCCHEM will move to a menu similar to the one shown in Figure 2-20. At this menu you may select the properties to be estimated by PCCHEM. You will not be able to provide measured values, however, since the responses provided at this menu must apply to all chemicals selected for the input group. When you are done with this menu, you will be returned to the PCCHEM Input Methods menu (Figure 2-18). At this point you may either run the program, or select either of the input options to provide information on additional chemicals to be included in the estimation run.

Display and save the output

The sequence of menus provided for this option is identical to that described in the

preceding section titled "Display Output to Screen."

NOTE: The PCCHEM Input Methods menu (shown in Figure 2-18) allows users to provide input in a variety of combinations. Users may also change output options between entry methods. The Output Options menu (illustrated in Figure 2-16) can be reached using the F9 key to back up from the Input Methods menu. Once you change the output option, the results of all subsequent chemical property estimates will be handled accordingly. Output for all chemicals input into PCCHEM prior to the change will be handled with the previous output mode.

- *Option 3. End of Input - Run*

When you are through entering structures for estimation, select this option. PCCHEM will present the program herald, followed by the estimation results if you chose to see them. Two output files are created if you are saving the output: CHEMXXX.OUT and CHEMXXX.DAT (XXX is replaced by a three-digit number that is sequentially assigned according to the number of datasets that have been created in that directory). The PCCHEM output file with the extension ".DAT" is meant to be used as input for the EXAMS and SESOIL models. The ".OUT" files contain summary output tables for each of the chemicals input by the user, and are designed to be printed if needed.

2.6 Hydrolysis of Water (PCHYDRO)

Hydrolysis is a chemical transformation process in which an organic molecule, RX, reacts with water, forming a new carbon-oxygen bond and cleaving a carbon-X bond in the original molecule. The net reaction is most commonly a direct displacement of X with OH to form ROH.

Hydrolysis is one of the most important reactions of organic compounds with water in aqueous environments and is a significant environmental fate process for many organic chemicals. It is actually not a single reaction but a family of reactions involving compound types as diverse as alkyl halides, carboxylic acid esters, carbamates, epoxides, organophosphonates, and nitriles.

The Structure Activity Relationships (SAR) equations used to estimate base-promoted hydrolysis have been developed by SRI International, Menlo Park, CA. These equations use Hammett and Taft parameters. The PCHYDRO program uses these equations to estimate hydrolysis rates for esters, carbamates, epoxides, isocyanates, amides, and acid chlorides. PCHYDRO is best suited to estimate hydrolysis of esters and carbamates since most of the other types of compounds hydrolyze too slowly or too quickly to make calculating the hydrolysis rate constants worthwhile. A compound is considered to hydrolyze too quickly if the hydrolysis occurs with a half-life of less than five minutes, and too slowly if the half-life is in the order of ten years. However, this should not deter you from entering any type of compound that you wish to estimate; PCGEMS will inform you if the hydrolysis rate is too slow or too fast.

The first menu you will see when you select this program is the output options menu. This menu is identical to the ones shown earlier for PCLOGP and PCCHEM (see Figure 2-16), and so will not be illustrated here.

Select the output option that you wish to use for this run. If you select either of the last two options, you will be required to enter an output label for the file in which PCFAP results will be stored. Choose a label which will help you identify file contents later. The name of the output file is automatically generated and contains an abbreviated version of the program name. When you have selected an option and entered a file label (if it is necessary), you will see the menu shown in Figure 2-24.

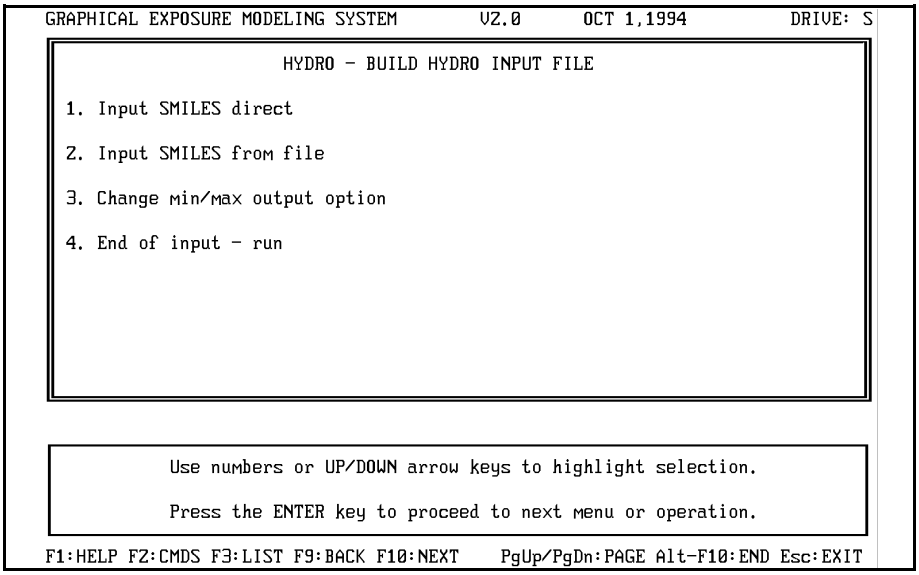

Figure 2-24. PCHYDRO Main Menu

You have four options in this navigational menu. Each option is explained below.

Option: 1. Input SMILES Direct

Users may directly enter the SMILES notations for the chemical of interest. The menu that appears when you select this option will not be illustrated, since an example has already been included in Figure 2-18. The SMILES notation method of structure entry is a simple process to learn. You must be familiar with the structure of the chemical for which you wish a hydrolysis half-life estimate. SMILES is also highly flexible in that there are several ways to specify a chemical structure. The rules for creating SMILES notation are explained in Subsection 2.3 of this chapter, and a reference is included for additional information. An alternate method for obtaining SMILES notation is the SMIGET program. The SMIGET program uses the chemical's CAS number to retrieve the SMILES notation from a database of about 19,000 chemicals. This program is explained in Subsection 2.2 of this chapter, and must be run before starting PCHYDRO. You can create an output file using SMIGET which can be used as an input file to PCHYDRO. The input file option is discussed below.

At this menu, you may also provide the chemical name. It will be helpful if you are planning to save the output, since it can be used later to identify the contents of the output file. When you have completed this menu, you will be returned to the menu shown in Figure 2-24. At this point, you may select this option again, or any of the other options, including running the program.

Option: 2. Input SMILES From File

When you select this option, a menu appears at which users may identify an input file name. Users do not need to provide the file extensions when identifying the input file. Users who have files that contain SMILES notation may use this file as input to the PCHYDRO program. This includes the output files from SMIGET, CLOGP, and PCCHEM. (Each of these programs is discussed in preceding subsections.) If you plan to use a file you created yourself, be aware that the file must be in the same format as those created by PCGEMS' estimation programs. There are three format restrictions:

- The file must be in ASCII format.
- The file extension must be ".DAT".
- The first variable in the file must contain the SMILES notation.

Once you enter an input file name, PCHYDRO will check to verify that the file exists and is in the proper format for use by the program. After this is done, you will see a menu similar to the one illustrated in Figure 2-7. PCHYDRO will automatically show the number of records included in the file. At this menu, users select the structures from the file which are to be included in the PCHYDRO estimation run.

You have the option of viewing all the contents of the file by typing the LIST command at either prompt. To view a portion of the file beginning with the first record, type LIST at the first prompt and a structure number at the second prompt. This will generate a list of structures starting with the first and ending with the structure identified in the second prompt. To view a portion of the file beginning with a user-selected structure and continuing to the end of the file, identify the structure number of the starting point at the first prompt and type the LIST command at the second prompt.

Once you have entered the beginning and ending numbers of the range of structures you wish to include in the estimation run, press ENTER and you will be returned to the menu shown in Figure 2-24. At this point you may continue to input structure information using either input option, you may change the output using the option described in the following paragraph, or you may run the program.

Option: 3. Change Min/Max Output Option

PCHYDRO output is defaulted to include only the results of an estimation run. This is referred to as minimum output. Users can expand the information included in the output by changing this parameter to maximum output. Maximum output includes two parameters used in the calculation of the hydrolysis rate: the E_s value, a measure of the steric effects, and the sigma value, a measure of the polar effects of the substituents of the chemical for which a half-life estimate is made. The maximum output option must be selected before an estimation run in order to include the parameter values in the PCHYDRO output. When you have changed the output option, the menu shown in Figure 2-24 will appear. At this point users may run the program, or enter additional structures using either of the available input methods, as described in the previous sections.

Option: 4. End of Input - Run

When you are through entering the structures for which you want hydrolysis estimates and you are satisfied with the type of output to be reported, select this option and press ENTER. PCHYDRO will take a moment to complete its calculations. You will then see the model results on your screen if you chose this option at the first menu, and you will see the name of the output file in the instruction section if you chose this option at the first menu. If the structure you entered was not an ester or a carbamate, the chemical may hydrolyze too quickly or too slowly. If this is the case for a chemical you entered, PCHYDRO will present a message advising you of this circumstance.

2.7 Fate of Atmospheric Pollutants (PCFAP)

Models such as ENPART can use, as part of their input, the rate of transport and transformation of a chemical in air. The dominant processes which control the concentration and mass of a chemical in the atmosphere are photolysis, oxidation by reaction with hydroxyl radicals, oxidation by ozone, and deposition on particles, in rain, and as dry molecules. The FAP model was designed by Heicklen Associates as a method for estimating the rate of atmospheric oxidation of a chemical by reaction with hydroxyl radical and ozone from structural information alone. In addition, if the solubility and vapor pressure of the chemical are known, the model can be used to estimate the rate of deposition of the chemical by adsorption on particles and dissolution in water droplets.

General Sciences Corporation adapted and modified the FAP model for inclusion in PCGEMS. PCFAP determines the chemical reaction rates by identifying structural fragments within the molecule which are susceptible to chemical reaction. Associated with each of these fragments are constants which are proportional to the chemical reaction rate. The total reaction rate for a chemical is obtained by summing the contribution to the reaction rate made by all of the identified fragments. The reaction rates calculated by PCFAP are second order. The PCFAP model performs calculations to convert the reaction rate to first order, for each of the following processes: abstraction of hydrogen, addition of hydroxyl radical, reaction with ozone, heterogeneous removal, surface lifetime, and lifetime in the troposphere which is 15 kilometers above the earth's surface. Greater detail on these processes is provided in the User's Guide to the GEMS FAP program (identical to PCFAP) [5].

To run PCFAP, enter its item number at the Chemical Property Estimation menu and press the ENTER key. The menu shown in Figure 2-25 appears. You must build an input file before you may run the PCFAP model. To do this, press the ENTER key when your cursor is over the first prompt.

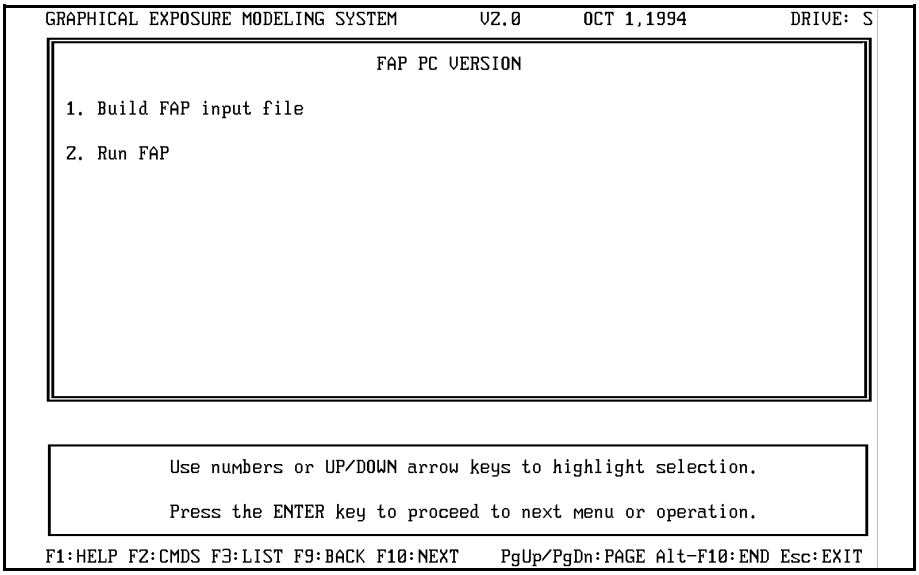

Figure 2-25. PCFAP Main Menu

The menu shown in Figure 2-26 appears when you select Option 1 from the PCFAP Main Menu.

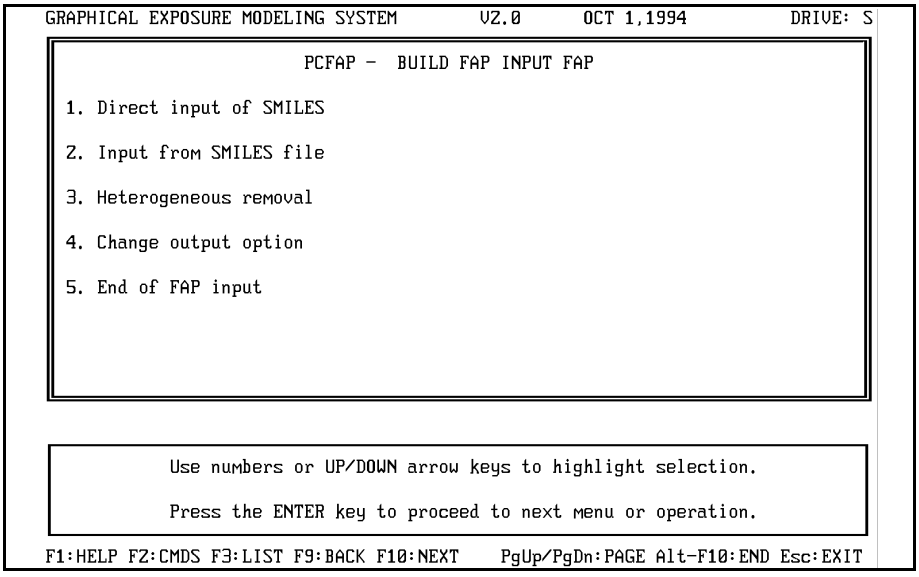

Figure 2-26. Build FAP Input File Menu

You must provide the SMILES notation of the chemical whose fate you wish to estimate. A brief explanation of the rules of SMILES notation and a reference is provided in Subsection 2.3. You may enter SMILES notations either directly or from a file already in existence, created through a run of another program such as SMIGET, or created outside PCGEMS and imported into the system.

Each option is explained below, along with the menus that appear when each option is selected.

Option: 1. Direct input of SMILES

The menu shown in Figure 2-27 appears when you select this option.

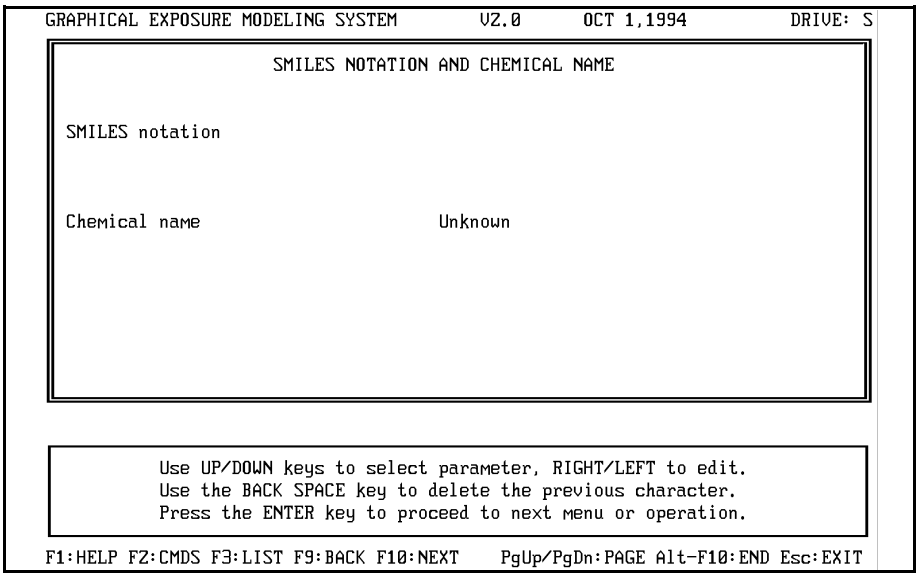

Figure 2-27. SMILES Notation - Direct Entry Menu

Enter the SMILES notation and the chemical name, if known. For information on the rules of SMILES notation, refer to Section 2.3. It will provide you with the basic rules about SMILES notation. Once you have entered the SMILES notation, you will be returned to the menu shown in Figure 2-26.

Option: 2. Input from SMILES File

To choose the file option, position your cursor over the prompt and press the ENTER key. The menu shown in Figure 2-28 appears in response below.

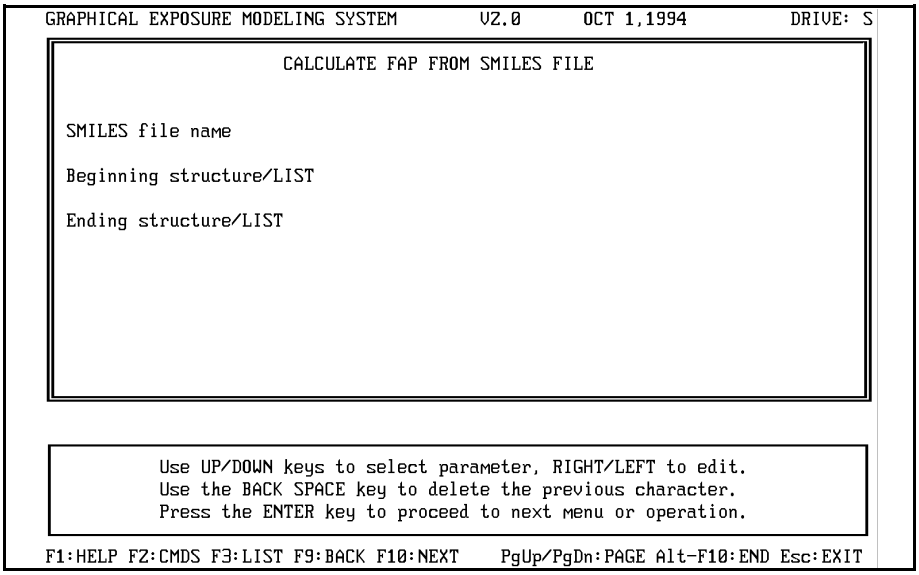

Figure 2-28. Calculate FAP from SMILES File Menu

This option assumes that you will be entering the output file from SMIGET as input here. If you wish to enter your own input file of SMILES notations, the file format for the file is provided below. Your user-entered file must follow the format provided below:

VARIABLE	VARIABLE TYPE
SMILES Notations	Alphanumeric (Maximum 120 characters)
Blank Space	
Chemical Name	Alphanumeric (Maximum 124 characters)
Keyword "CASNO"	5 characters
Blank Space	
CAS Numbers	Numeric (Maximum 13 characters)

An example of what your file should look like is provided below:

```
c(cccc1)(c1)-CC Benzene, ethyl- CASNO 100414
c(cccc1)(c1)-C Benzene, methyl- CASNO 108883
c(=CCl)(Cl)Cl Ethene, trichloro- CASNO 79016
```

Once you have entered a file name, you can review a list of the SMILES notations in your file by entering a LIST at the enter the second or third prompt. You may estimate properties for more than one chemical at a single session by entering the starting and ending numbers of a range of structures in your file that you wish to analyze. Then press the ENTER key and the Build PCFAP Input File (Figure 2-26) reappears, indicating that your input file has been created. (The "LIST" command has been described as part of the SMIGET discussion, found in Subsection 2.2 of this chapter.)

Option: 3. Heterogenous Removal

The heterogeneous removal of a chemical is calculated by a routine that requires the user to enter the normal vapor pressure and water solubility of the compound. The normal default value for this prompt is "N". Only if you have the property rates required, should you change this option to "Y". You may use PCCHEM to estimate the water solubility and vapor pressure **before** using this program. To do that from this point in PCGEMS, enter the ALT/F10 (END) command and proceed to the PCCHEM program. When you have obtained the property estimates, you may return to this program and estimate heterogeneous removal.

Option: 4. Change Output Option

You may also choose to have the calculations included as part of the program output. The default value for this option is "minimum." To change it to maximum, move your cursor to the "Change Output" option and press ENTER; then change the value of the prompt to "Y".

Option: 5. End of FAP Input

When you have entered the information necessary to build an input file, move your cursor to the last prompt, "End of PCFAP input," and press the ENTER key to proceed.

The main PCHYDRO menu shown in Figure 2-25 reappears. Position your cursor over the second prompt, "Run PCFAP," and press ENTER. The menu shown in Figure 2-29 appears.

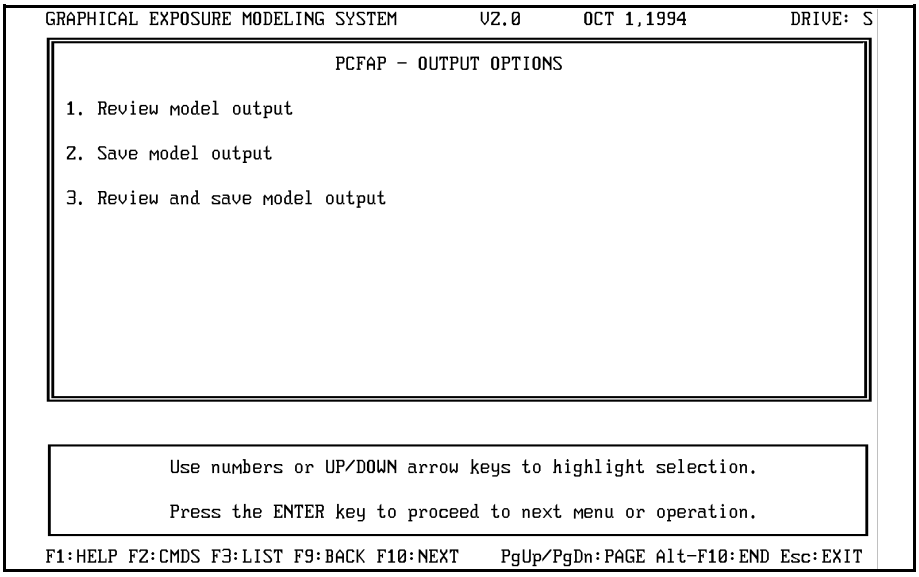

Figure 2-29. PCFAP Output Options Menu

As you can see, you have three options:

Option: 1. Review Model Output

You may choose to view the model output on your PC screen and not save it.

Option: 2. Save Model Output

You may choose to save model output to a file and not view it on the screen.

Option: 3. Review and Save Model Output

You may choose to both display the model results on screen and save them.

If you do choose to save model results, a menu will appear asking you to enter a label for the output file. Provide a label that will help you identify the file contents later. The model results are displayed page by page if you chose to review the results. You will be queried at the end of each page whether or not you wish to see the rest of the output results.

2.8 Draw SMILES Notation (DRAWSMI)

This program allows you to draw the structure of a chemical or a list of chemicals either by entering the SMILES notation for the chemical you wish to draw or by entering the CAS numbers. If you are entering the CAS number of the chemical for which you would like the structure, please be aware that you must have the CAS Number/SMILES notation database available in the PCGEMS data path, and the CAS number that you enter must be in the database for this program to work.
To select this program, press ENTER when the cursor is over this program item in the Estimation Menu or enter the item number and press ENTER. The menu shown in Figure 2-30 appears.

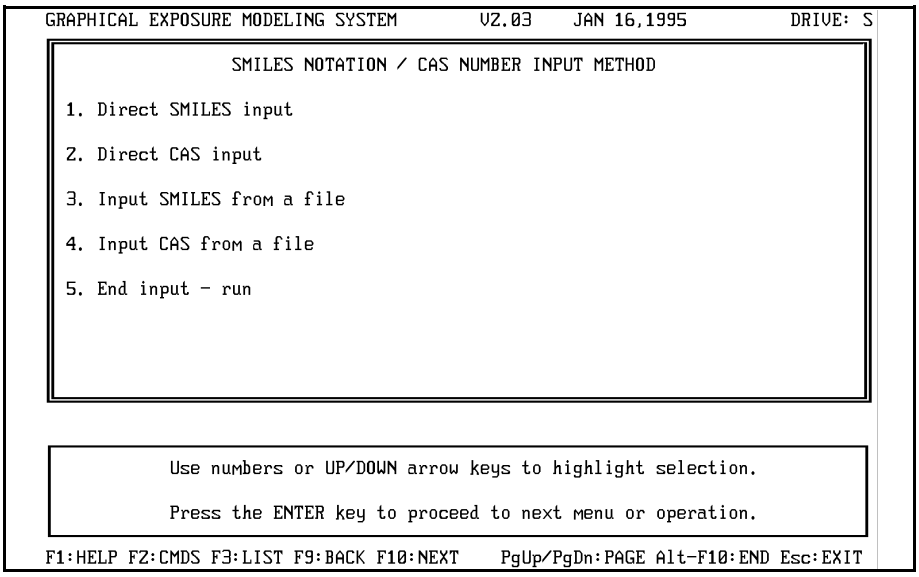

Figure 2-30. DRAWSMI Main Menu

You have six options in this menu. Each option is explained below:

Option 1: Direct SMILES Input

You may enter the SMILES notations directly. The rules for using SMILES notation to depict a chemical are provided in the help message that is associated with the SMILES entry. To access the help message, press the F1 key at the prompt where you are asked to enter the SMILES notation. You may also choose to enter the CAS Number and the chemical, if you know it. This information will appear along the drawn structure.

Option 2. Direct CAS Input

In order to use this option, you must have the CAS Number/SMILES database and the location of this database must have been entered in the PCGEMS file configuration. The chemical will only be drawn if the chemical whose CAS Number you have entered can be found in the database. If you enter a CAS number that is not in the database, you will be informed of this immediately after you enter the CAS number. When you have entered the CAS number, press ENTER and you will be returned to the Main DRAWSMI menu, where you may choose to enter another

chemical or run the program.

Option 3. Input SMILES from a File

This option allows users to specify a file containing SMILES notations for use as input for the program. If any PCLOGP output files exist in your default directory, the latest CLOGP*.DAT will appear as the default input file. You may also enter a SMILE*.DAT output file created by SMIGET here, or a SMILE*.DAT output file created by DRAWSMI on a previous occasion. The F3 key will provide you with a list of all the CLOGP*.DAT file that exist in your catalog manager file. For other file names, you will have to exit the DRAWSMI program and then call the F3 key.

The input SMILES file can also be through user-entry outside PCGEMS. If you are using a file that you have created yourself, the file format for the file is provided below. The file must have the format provided below:

FILE FORMAT

The input file must be in ASCII format. The SMILES notations must be in the first column of the file. There must be three columns that are separated by a blank space. The format of this file is provided below:

VARIABLE	VARIABLE TYPE
SMILES Notations	Alphanumeric (Maximum 120 characters)
Blank Space	
Chemical Name	Alphanumeric (Maximum 124 characters)
Keyword "CASNO"	5 characters
Blank Space	
CAS Numbers	Numeric (Maximum 13 characters)

An example of what your file should look like is provided below:

```
c(cccc1)(c1)CC Benzene, ethyl- CASNO 100414
c(cccc1)(c1)-C Benzene, methyl- CASNO 108883
C(=CCl)(Cl)Cl Ethene, trichloro- CASNO 79016
```

Option 4. Input CAS from a file

You may choose to enter a user-entered file containing CAS numbers. The file format for the file is provided below:

FILE FORMAT

- The file must be in ASCII format.
- The CAS numbers must be the first variable in each record in the file and each CAS number can only be a maximum of 13 characters long.

Option 5. End Input - Run

When you have entered all the chemicals that you want, select this option. You will see a message informing you that a batch file is being created followed by the menu shown below in Figure 2-31.

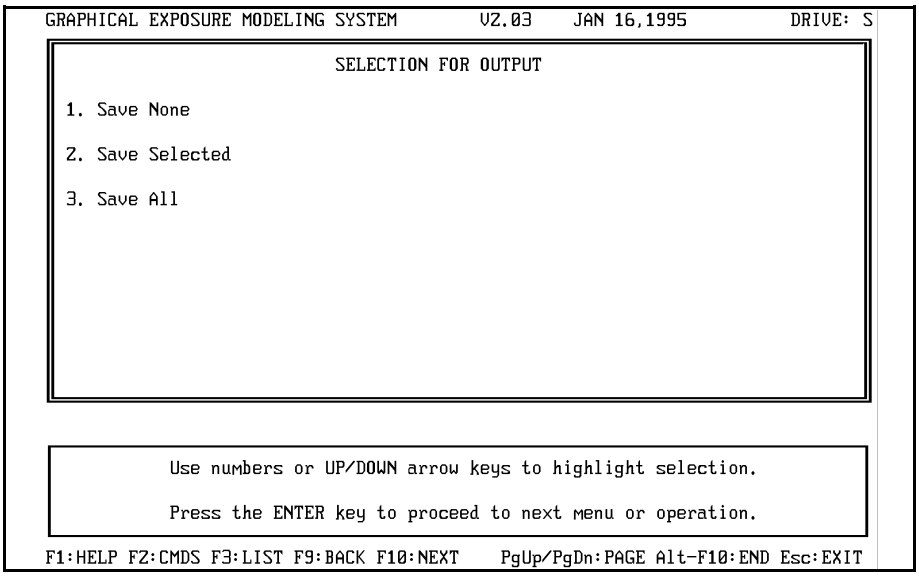

Figure 2-31. DRAWSMI Output Options

Please note that the save option does not save the graphic images of the chemicals that are to be drawn but rather the SMILES notations of these chemicals. The saved chemicals are stored in a file called SMILE*.DAT, which may be used as input for the chemical property estimation programs: PCLOGP, PCCHEM, PCHYDRO, and PCFAP. It also may be used as input to DRAWSMI using option 3 from the DRAWSMI main menu.

The first option will not save any of the chemicals drawn and the last option will save all the chemicals that are drawn. The second option will ask you after every drawn chemical whether you wish to save that specific chemical in a file. Select an option and press ENTER. The SMILES structure will be drawn on the screen. An example of this is provided below.

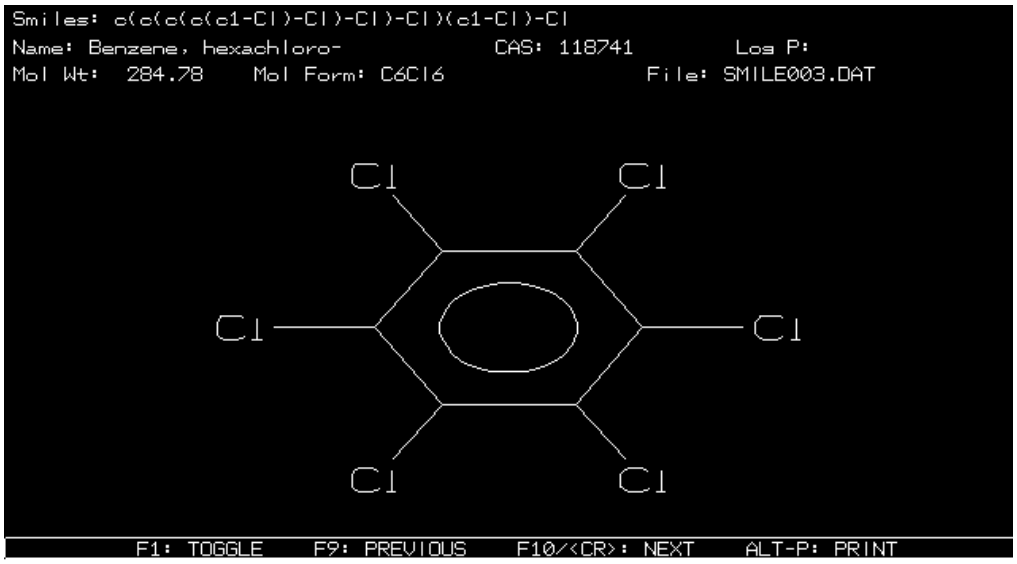

Figure 2-32. Example of DRAWSMI Output

There will be a toggle bar that appears on the bottom of your screen. The print option on this toggle bar will only print to a very limited set of dot-matrix printers. Refer to Chapter 1 if you have questions on printing. When all the chemicals are drawn, you will see a message providing you with the file name of this file if you chose to save it (SMILE*.DAT). The file can be used as input file for the other property estimation programs.

2.9 Run User-Installed Estimation Programs

Use this capability to access any of your own estimation programs. Before being able to use such a program, however, you must install the program using PCGEMS's Utilities function. To install a non-PCGEMS estimation program under PCGEMS, you must do two things:

Step 1.

Install the program using the XINSTALL function of PCGEMS's utilities. To do so,

proceed to the Utilities operation and select option 1, "Install a Non-PCGEMS Program". Refer to Subsection 7.3 for more information on this utility. Keep in mind that once you have installed a program under this option, the program name and path will always be listed under this option. You cannot remove this listing. However, if PCGEMS does not find the program where you indicated it would be, the listing will be deleted automatically.

Step 2.

Install the XRUNNPC program. If you have installed your program correctly, the information that you entered identifying the program when installing it should appear on the screen when you select this option. When the cursor is over the program name, press ENTER to access this program.

Once you have carried out both steps, you may run the program by selecting "Run User-Installed Estimation Programs" at all subsequent sessions.

Contents

Figures

Tables

Index